DOCUMENT RESUME

ED 384 340 IR 017 203

AUTHOR Diepold, Peter, Ed.; Rusch-Feja, Diann, Ed.
TITLE Elektronische Informationsdienste im Bildungswesen (Electronic Information Services in Education) Gesellschaft Information Bildung Conference (GIB) (2nd, Berlin, Germany, November 17-18, 1994).
REPORT NO ISBN-3-9804450-0-3; ISSN-0945-9898
PUB DATE 95
NOTE 204p.
PUB TYPE Collected Works - Conference Proceedings (021)
LANGUAGE German

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS Computer Software Selection; Educational Quality; *Educational Technology; Evaluation Methods; Foreign Countries; *Information Networks; *Information Services; Optical Data Disks; Teaching Methods
IDENTIFIERS Internet; West Germany (Berlin)

ABSTRACT

These papers on educational technology were presented in three workshops at the second annual conference of the Society of Information Education (GIB). Discussion includes electronic networks, CD-ROMs, and online databases in education, the quality of educational software, database services and instructional methods, and the use of the Internet in supporting research and teaching. An evaluation of the German educational literature CD-ROM, "Literaturdoijumentation CD-Bildung," is included. Specifically, the themes of the workshops were presented as follows: (1) production of electronic services, with reports on the production of databases, CD-ROMs, hypertext systems and a full-text CD-ROM; (2) evaluating database quality for content, organization and user interface with regard to a thesaurus; and (3) criteria for the selection of literature administration software and a comparative study of several popular literature administrative programs. Contains 76 references. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 384 340

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Rusch-Feja

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR017203

BEST COPY AVAILABLE

Peter Diepold / Diann Rusch-Feja

(Herausgeber)

Elektronische Informationsdienste im Bildungswesen

2. GIB-Fachtagung

17. - 18. November 1994 in Berlin

GIB

Gesellschaft Information Bildung

Berlin 1995

Schriftenreihe der Gesellschaft Information Bildung, 2

ISSN 0945-9898

Elektronische Informationsdienste im Bildungswesen ;
2. GIB-Fachtagung Berlin, 17.-18.11.1994 /
Peter Diepold, Diann Rusch-Feja (Hrsg.). -
Berlin : Gesellschaft Information Bildung, 1995.

(Schriftenreihe der Gesellschaft Information Bildung; 2)
ISSN 0945-9898
ISBN 3-9804450-0-3
NE: Diepold, Peter [Hrsg.]; Rusch-Feja, Diann [Hrsg.]; GT

Bezug über:

Gesellschaft Information Bildung (GIB)
Geschäftsstelle:
Arnimallee 10
D-14195 Berlin
Tel.: (030) 838 5888
Fax: (030) 838 5889

ISSN 0945-9898
ISBN 3-9804450-0-3

© Gesellschaft Information Bildung und Autoren

Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb des Urheberrechtsgesetzes bedarf der Zustimmung der Gesellschaft Information Bildung und des Autors. Das gilt insbesondere für Vervielfältigungen jeder Art, Übersetzungen, Mikroverfilmungen und für die Einspeicherung in elektronische Systeme.

Druck: Wittchen, 37176 Nörten-Hardenberg

Inhaltsverzeichnis

Peter Diepold / Diann Rusch-Feja
Einleitung 1

Hauptvorträge

Wolfgang Meyer
Elektronische Informationsdienste in Schulen 3

Peter Diepold
Elektronische Informationsdienste für Wissenschaftler 15

Robert Funk
Zusammenfassung der Diskussion 33

Friedrich Rost
Erziehungswissenschftlich relevante elektronische
Dienstleistungsangebote: Ein Überblick 37

Alexander Botte
Die "Literaturdokumentation Bildung auf CD-ROM"
des FIS Bildung
Eine erste Bilanz nach acht Monaten 69

Rosemarie Kohls
Zusammenfassung der Diskussion 81

Workshops

Workshop 1:
Erfahrungen in der Produktion von elektronischen Diensten

Rainer Dechsling
Versuch einer Typologie der textorientierten
elektronischen Fachinformation
Datenbank, Hypertext oder linearer Text? 83

Sylvia Springer
Erfahrungen in der Produktion von elektronischen Diensten
Die Pestalozzi-Volltextdatenbank 91

Doris Bambey
Vorbereitung und Entwicklung der CD-ROM Bildung
Probleme, Methoden und Instrumente 125

Zusammenfassung: Alexander Botte 133

Workshop 2: Bewertung der Qualität von Datenbanken

Heinz Ziegler
Einige Anforderungen an Recherchesoftware
bei Verwendung eines Thesaurus 135

Yrla Labouvie / Petra Konz
Qualitätskriterien und Qualitätssicherungsmaßnahmen
aufgezeigt an der Datenbank PSYNDEX 157

Walther Umstatter
Bewertung der Qualität von Datenbanken 169

Zusammenfassung: Diann Rusch-Feja 177

Workshop 3: Software-Entscheidungen
Kriterien zur Auswahl von Literaturverwaltungs-Programmen

Astrid Diening
Literaturverwaltungssysteme im Vergleich 181
Willi Reinecke
Zur Auswahl von Softwareprodukten
für das Bibliotheks- und das Informationswesen 193

Zusammenfassung: Bernd Christopher 203

Teilnehmerverzeichnis 205

Elektronische Informationsdienste im Bildungswesen

Eine Einführung

Am 17. und 18. November 1994 fand an der Humboldt-Universität zu Berlin und am Max-Planck-Institut für Bildungsforschung die zweite Fachtagung der Gesellschaft Information Bildung (GIB) statt. Nachdem ein Jahr zuvor die GIB auf ihrer ersten Fachtagung in Frankfurt am Main eine generelle Standortbestimmung der Bildungsdokumentation in Deutschland vorgenommen hatte,[1] war das Ziel der zweiten Fachtagung, die Nutzung *elektronischer Dienste* für Pädagogik, Erziehungswissenschaft und Bibliotheken vorzustellen, zu diskutieren und Konsequenzen für die Bildungsdokumentation zu ziehen.

Unter den fast 100 Teilnehmern der Berliner Tagung waren Hersteller von Fachinformationen (Dokumentare, Verleger, Medienspezialisten), Informationsvermittler (Bibliothekare) sowie Erziehungs- und Bibliothekswissenschaftler. Dies entspricht den Zielen der 1991 gegründeten GIB, Anbieter *und* Nutzer von Information im Bildungsbereich zusammenzubringen.

Das gewählte Tagungsthema zeichnet sich durch hohe Aktualität, nicht geringe Brisanz und erhebliche Konsequenzen für die Zukunft der Bildungsdokumentation und ihrer Berufe aus. Ein hochentwickeltes Land wie Deutschland ist auf ein funktionierendes Bildungssystem angewiesen, das seinen Absolventen die für die Zukunft nötigen Qualifikationen vermitteln kann. Für Bildung stehen aber zunehmend weniger Ressourcen zur Verfügung, dabei steigen die Kosten für traditionelle Bildungsmedien, Bücher und Zeitschriften. Doch entwickelt die Informationsgesellschaft auch Möglichkeiten der kostengünstigen Herstellung, Verarbeitung, Verbreitung und Nutzung von Informationen, die es dem Wissenschaftler, dem Bildungsadministrator, dem Lehrer oder dem Ausbilder ermöglichen, von seinem Arbeitsplatz aus über

[1] Diese Tagung ist dokumentiert in der Schriftenreihe der GIB, Bd. 1: P. Diepold / D. Rusch-Feja (Hrsg.): Bildungsdokumentation Heute. Fachinformation für Wissenschaft und Praxis. Berlin 1994.

den vernetzten Computer nicht nur auf elektronische Kataloge, Datenbanken und Volltexte direkt zuzugreifen, sondern selbst Informationen anzubieten.

Das bleibt nicht ohne Konsequenzen für traditionelle Arbeit. Unter dem Aspekt weltweiter Vernetzung verwischen sich die Funktionen des Bibliothekars und des Dokumentars ebenso wie die des Wissenschaftlers und des Verlegers.

Internet, Online Public Access Catalogs, CD-ROM-Recherchen am Arbeitsplatz, virtuelle Bibliotheken oder Electronic Publishing sind einige der Schlagworte, die diese Entwicklung thematisieren. Zu den ungelösten Problemen elektronischer Informationsdienste gehören Themen wie Korrektheit und Authentizität von Informationen, das Recht auf Verschlüsselung von persönlichen Informationen gegenüber dem Versuch staatlicher Kontrolle, Datenschutz und Datensicherheit, neue Möglichkeiten der Indexierung, Qualität der Erschließung, Copyright, Kosten schneller Dokumentenlieferung u. a. m.

Auf der Tagung wurden Pilotprojekte und Innovationen präsentiert, Versuche zur Präzisierung und Kategorisierung von Qualitätsmerkmalen vorgestellt sowie Diskussionen über Kosten, Zweck und Nutzungen geführt. Sie sind auf den folgenden Seiten dokumentiert.

Während des Rahmenprogramms bestand die Möglichkeit, die Bibliotheken, CD-ROM-Netze, Volltextdatenbanken und Informationsdienste der veranstaltenden Berliner Institutionen praktisch kennzulernen.

Die GIB dankt der Humboldt-Universität, der Bibliothek für Bildungsgeschichtliche Forschung und dem Max-Planck-Institut für Bildungsforschung für ihre Gastfreundlichkeit und deren Mitarbeitern für tatkräftige Unterstützung bei der Durchführung der Fachtagung. Wir danken den Autoren und den Berichterstattern für die zusätzliche Arbeit, die sie sich mit der Veröffentlichung ihrer Beiträge unterzogen haben, und Bärbel Bäcker von der Abteilung Pädagogik und Informatik der Humboldt-Universität für das Lesen der Korrekturen und die Herstellung der Druckvorlage.

Berlin, im Mai 1995

Peter Diepold *Diann Rusch-Feja*

Elektronische Informationsdienste in Schulen

Wolfgang Meyer

Deutschland hat auf dem Gebiet der schulischen Nutzung elektronischer Netzwerke im Vergleich zu anderen Industrieländern einen deutlichen Nachholbedarf. Während in Großbritannien ca. 4000 Schulen allein das von BT (British Telecom) betriebene Schulnetzwerk CAMPUS 2000 nutzen, ist in Deutschland die Zahl der vernetzten Schulen noch im Bereich der Hunderter.

Computer- und Telekommunikationsnetze stellen heute unabhängig von Entfernungen und lokalen Rahmenbedingungen Bildungsangebote bereit und vernetzen Lerngruppen. Mailbox-Systeme bieten die schnelle Übermittlung von Text-, Bild-, Ton- und Video- und Programmdateien sowie den Zugang zu Computer-Konferenzen und Datenbanken.

1 Elektronische Post

Elektronische Post erlaubt es, papierlos Briefe, Aufzeichnungen, Informationen zu versenden. Die Texte entstehen auf einem Computer und erreichen den Empfänger über das Telefonnetz. Zwischenstation ist ein Mailbox-Rechner. Email ist schnell. Sobald ein elektronischer Brief fertiggestellt ist, kann er abgeschickt und Augenblicke bis wenige Stunden später von seinem Empfänger gelesen werden. Elektronische Post kann zu jeder Tages- oder Nachtzeit abgeschickt bzw. empfangen werden, unabhängig von internationalen Zeitzonen. Der Zugang zu einem Mailbox-System kann über jeden Telefonanschluß erfolgen. Professionelle Systeme sind zuverlässig. Auf Wunsch erhält der Absender automatisch eine Empfangsbestätigung und weiß damit genau, ob und wann sein Brief gelesen wurde. Verteilerlisten erleichtern die Handhabung von "Massensendungen": Wenn die Namen und Kennungen der Adressaten bekannt sind, kann eine Mitteilung gleichzeitig an eine Gruppe von Empfängern geschickt werden. Der zusätzliche Zeitaufwand beträgt nur Sekunden, wenn der Brief statt an einen an hundert Empfänger geschickt wird.

Im Schuljahr 1991/92 konnten über 100 Schulen in verschiedenen Bundesländern das britische Schulnetz CAMPUS 2000 erproben. Sie kommunizierten überwiegend mit britischen, aber auch mit skandinavischen, australischen, amerikanischen und kanadischen Partnern. Viele beteiligten sich an Projekten, überwiegend im Bereich des Fremdsprachenunterrichts.

Es hat sich gezeigt, daß die als elektronische Post in Projekten ausgetauschten Daten - Arbeitsberichte, Referate, landeskundliche Informationen, persönliche Mitteilungen und Rückfragen - eine Bereicherung des Unterrichts sind. Sie unterscheiden sich nicht nur durch die Aktualität von didaktisch aufbereiteten Lehrbuchtexten, sondern vor allem durch Authentizität.

> "The notion of students being the providers, as well as the consumers, of information available online is an important one for citizens of the Information Age." [1]

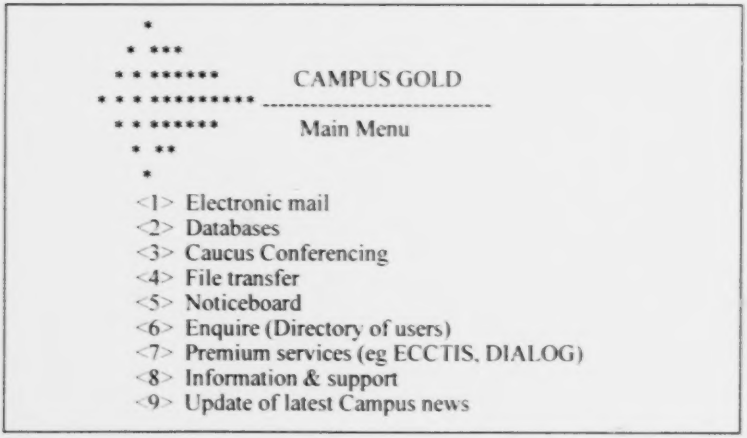


Bild 1: CAMPUS 2000 Main Menue

2 Computer Conferencing

Computer-Konferenzen sind orts- und zeitunabhängige Diskussionsforen. Die Teilnehmer treffen sich nicht persönlich. Diskussionsbeiträge können als elektronische Post von jedem Ort und zu jeder Zeit geliefert werden. Jeder

[1] Judy Harris: Making the Hypertext Happen, in: *The Computing Teacher* December/January 1994/95, S. 36-41 (Email: jbharris@tenet.edu)

Teilnehmer kann sich einen Überblick über den gesamten Diskussionsstand verschaffen. Die Konferenz ist dabei das kollektive Gedächtnis der Gruppe.

Computer Conferencing wird seit vielen Jahren genutzt. Im RAPPI-Projekt (Reseau d'ateliers pédagogique pilote) experimentierten bereits 1985 ca. 50 Schulen aus Kanada, Frankreich, Italien und Großbritannien mit dem neuen Medium.[2]

Das wichtigste Kriterium zur Unterscheidung von Konferenzen ist die Regelung der Zugangsberechtigung:

- Offene Konferenzen können von allen Teilnehmern eines Netzwerkes genutzt werden;
- geschlossene Konferenzen sind einem eingeschränkten Teilnehmerkreis vorbehalten;
- in anonymen Konferenzen ist nicht erkennbar, von wem einzelne Beiträge stammen. Sie erfreuen sich als neue Spielart des betrieblichen Vorschlagswesens in den USA großer Beliebtheit.

Anwendungsmöglichkeiten für Computer-Konferenzen im schulischen Kontext:

- Austausch von Informationen und Erfahrungen, insbesondere im Rahmen von Projekten;
- Ausarbeitung von Berichten, Materialien etc. durch eine Gruppe von Menschen, die an verschiedenen Orten wohnen;
- Koordination von Projekten auf nationaler und internationaler Ebene;
- interaktive Durchführung von Fernkursen: die Teilnehmer haben Kontakt zu allen anderen Kursteilnehmern und zum Tutor.

Beispiel: World Tour

1992 wurde das Projekt World Tour als CAUCUS-Konferenz durchgeführt. 100 Klassen gingen auf eine virtuelle Weltreise und besuchten ihre Mitschüler in zwölf ausgewählten Schulen in Großbritannien, Irland, USA, Australien, Japan, Belgien, Rußland, Schweden und Ostfriesland. Für jeweils eine Woche war eine der zwölf Schulen Gastgeber für die Weltreisenden, mußte für einen angenehmen "Aufenthalt" der Gäste sorgen und deren Informationsbedürfnis als virtuelle Touristen befriedigen. Die Norder Schülerinnen und Schüler schrieben zunächst eine Kurzinformation über die Stadt. Es folgten Beschreibungen ostfriesischer Sitten und Gebräuche, des lokalen Freizeitangebotes und der Anreise vom nächsten Flughafen.

[2] Vgl. Maria Cioni & Associates Inc.: An Evaluation Report of the RAPPI Project. Canadian Department of Communication, 1986. Einen guten Überblick bietet A.R.Kaye (Hrsg): Collaborative Learning through Computer Conferencing. The Najaden Papers. Berlin/Heidelberg (Springer), 1992.

3 On-Line-Datenbanken

Wir unterscheiden zwischen Datenbanken für Schulen, die im Kontext von Modellvorhaben oder durch Unterstützung aufgeschlossener Schulbehörden und Universitäten angeboten werden, und kommerziellen Datenbanken, deren Nutzung im Unterricht sinnvoll erscheint.

3.1 Von Schulen und für Schulen entwickelte Datenbanken

Das gemeinsame Problem aller bisherigen Versuche auf diesem Gebiet ist das der kritischen Masse. Dies betrifft sowohl den Umfang des zusammengetragenen Materials als auch der aktiven Nutzer. Statt eine Reihe erfolgloser Bemühungen aufzuzählen, beschränken wir uns auf ein herausragendes Beispiel, die Hampshire Database bei CAMPUS 2000.

In den kommenden Jahren wird auch für deutsche Schulen das INTERNET an Bedeutung gewinnen. Zur Zeit ist ein vollwertiger INTERNET-Zugang für deutsche Schulen, der über den Austausch von elektronischer Post und das Abonnieren von Foren hinausgeht, mit zu hohen Kosten verbunden.

3.1.1 Hampshire Local Database

HAMPSHIRE EDUCATION CAMPUS 2000
LOCAL DATABASE

1 Database Help / Commands
2 Geography GCSE Pupils' Work
3 IT Register - Support Team
4 TELETEXT NEWS in text format
5 Marwell Zoo Database
6 Micronews (back numbers)
7 Software Reviews
8 School Library Database
9 Intercultural Education
11 Modern Language Database
12 COMMS - Communications
13 Needs of the Future Group
14 IT INSET Course Details
15 News in German + SKY
16 Flexible Learning
17 Contact System Manager
18 Italia Novanta
10 Places to Visit in Hampshire

Die interessanteste Abteilung verbirgt sich hinter dem Menüpunkt 4. Hier können aktuelle Nachrichten des Tages abgerufen werden. Es sind täglich mehrseitige Meldungen, die aus dem britischen Videotext, dem französischen Minitel und neuerdings auch aus dem deutschen Fernsehen stammen. Wer diese Meldungen täglich abruft und auf der Festplatte ablegt, hat in kurzer Zeit einen Fundus von Texten, der als Mini-Datenbank genutzt werden kann. Die Suchfunktion der Textverarbeitung kann eingesetzt werden, um Namen oder Begriffe aufzufinden.

3.2 Unterrichtsrelevante kommerzielle Datenbanken

Der Modellversuch MODIS in Nordrhein-Westfalen hat gezeigt, daß Pressearchive in besonderer Weise als Informationsquelle für die Arbeit in unterschiedlichen Fächern geeignet sind, wobei auf ergänzende Recherchen in Fachdatenbanken weitgehend verzichtet werden kann.

Die Gebührenstruktur der kommerziellen Dienste hält in der Regel schulische Nutzer fern. Eine Reihe von Datenbanken ist für Schulen schlicht zu teuer. Die Verlagsgruppe Handelsblatt bietet über Datex-J rund zwanzig Pressedatenbanken an, darunter *AFP, Berliner Zeitung, Der Tagesspiegel, Handelsblatt, Süddeutsche Zeitung* und *Wirtschaftswoche*. Die Anschaltgebühr beträgt DM 36,00 pro Stunde und im Schnitt DM 3,50 pro Dokument. Bei der geringen Geschwindigkeit von Datex-J laufen bei ernsthafter Nutzung schnell hohe Gebühren auf.

3.2.1 PROFILE EDUCATION

Die deutschen Datenbankanbieter offerieren bisher keine Sonderkonditionen für Schulen. In Großbritannien und den USA ist die Bereitschaft der kommerziellen Anbieter, Schulen die Benutzung ihrer Dienste zu finanzierbaren Preisen zu ermöglichen, deutlich größer.

CAMPUS 2000 zum Beispiel hat ein Gateway zu PROFILE EDUCATION mit *Financial Times, Times & Sunday Times, Guardian, Daily Telegraph, Independent, New Scientist* und *Today*. Die Bestände reichen bis in die Mitte der achtziger Jahre zurück. Recherchen sind schnell und können auch in mehreren Zeitungen gleichzeitig durchgeführt werden. Die Kosten liegen bei knapp DM 90,00 pro Stunde. Zusätzliche Kosten für Recherchen und angezeigte Dokumente entstehen nicht.

3.2.2 Das Bildungsangebot von Dialog

DIALOG CLASSMATE ist eine Auswahl von 85 aus dem von DIALOG Information Services angebotenen 450 Datenbanken von einem der größten amerikanischen Datenbankanbieter. Bildungseinrichtungen können CLASSMATE zu einem Bruchteil des Preises für kommerzielle Kunden nutzen. CAMPUS berechnet einen Aufpreis von ca. DM 36 pro Stunde. Das Angebot reicht von Agriculture bis Travel und deckt viele Bereiche ab, die für Schule und Ausbildung von Interesse sind. Der weniger geübte Benutzer kann sich über Menüs führen lassen; der Profi kommt mit direkten Befehlen schneller zum Ziel. Das Menü zeigt die thematische Gliederung der Datenbanken:

1 Agriculture & Nutrition
2 Bibliography-Books & Monographs
3 Business Information
4 Chemistry
5 Computers & Electronics
6 Directories & Reference
7 Education
8 Law & Government
9 Medicine, Biosciences,
10 News & Current Affairs
11 Popular Information
12 Science & Technology
13 Social Sciences Humanities
14 Official Airline Guide

Das Untermenü News & Current Affairs läßt erkennen, welche riesigen Textbestände durchsucht werden können:

News & Current Affairs
U.S. Regional Newspapers (Full Text)
1 Boston Globe
2 Chicago Tribune
3 Los Angeles Times
4 Philadelphia Inquirer
5 San Francisco Chronicle
6 San Jose Mercury News
7 Washington Post Online
Newspapers (Abstracts & Summaries)
8 National Newspaper Index (tm)
9 Newsearch
10 USA Today Decisionline
Financial News (Abstracts & Summaries)
11 Standard & Poor's News 7/85 - present
12 Standard & Poor's News 6/79 - 6/85
13 Canadian Business & Current Affairs
Current Affairs (Abstracts & Summaries)
14 Facts on File
15 Current Digest of the Soviet Press
16 Magazine Index (tm)
17 Canadian Business & Current Affairs

Newswires (Full Text)
18 UPI News (most current 6 months)
19 UPI News (1983 forward)

Deutsche Schulen können DIALOG Classmate zu einem günstigen Preis auch direkt abonnieren. Der Zugang erfolgt über das Datennetz GNS Tymnet, das in Deutschland eine Reihe von Einwählknoten unterhält. Einschließlich der Netzwerkgebühren kostet eine Stunde Datenbankrecherche 15 USD.

Handbücher, Arbeitsbücher, Lehrerhandbücher und eine Videokassette mit einem "Qick-Start" Selbstlernprogramm erleichtern die Durchführung eines Unterrichtsprojekts Datenbankrecherche.[3]

3.2.3 CompuServe

CompuServe ist ein rasch wachsender amerikanischer Online-Dienst mit rund 2,5 Mio Teilnehmern. Er bietet elektronische Post und eine große Anzahl von Foren und kommerziellen Informationsdiensten, insgesamt etwa 2000 Datenbanken. 700 Fachpublikationen und Zeitschriften liegen im Volltext vor, darunter *Daily Telegraph, Financial Times, Observer, Economist, European, Washington Post*. Durch die 10 Knotenrechner in Deutschland sind die Kosten für die Basisdienste moderat; die Monatsrechnung des Autors (für Email und die gelegentliche Nutzung von Foren) liegt bei etwa 15 Mark.

Ein für Schulen sicher interessanter Ort im CompuServe Cyberspace ist das SPIEGEL-Forum, das nach seiner Einführung im Frühjahr 1994 bis zum Ende des Jahres über 20.000 Datenreisende aufgesucht haben. Es enthält ausgewählte Artikel, Grafiken und Fotos. Die verantwortlichen Redakteure schätzen die Möglichkeit der direkten Diskussion mit den Lesern.

Am 20.12.1994 stellte sich der Sächsische Ministerpräsident als erster deutscher Politiker den Fragen der Teilnehmer einer Online-Konferenz im SPIEGEL-Forum. Das *Wall Street Journal Europe* berichtete am 22.12.94 unter folgender Überschrift:

> Meeting in Cyberspace Helps Germans Edge Onto the Infobahn
> Governor of Saxony Fields Questions, Gets a Lesson On Baud and Megabyte
> The 64-year-old governor of the eastern German state of Saxony is no Al Gore, the U.S. vice president who has claimed a front seat in the bandwagon rolling down the infobahn. Techno-novice though he is,

[3] Kontakt: Data-Star DIALOG Europe, Ostbahnhofstraße 13, 60314 Fankfurt am Main, Tel. 069-444063 Fax 069-442084. GNS Tymnet, BT Deutschland GmbH, Eschborn/TS, Tel. (06196) 96040

> however, Mr. Biedenkopf is believed to be the Continent's first leading politician to field inquiries from the public in a town-meeting in cyberspace.[4]

Auf die Frage, ob Schulen einen kostenlosen Zugang zum Internet erhalten könnten, antwortete Biedenkopf: "Soweit der Zugang zu Internet für den Schulunterricht notwendig ist, muß ihn der Staat als Schulträger auch finanzieren." Und an anderer Stelle: "Das müssen die Pädagogen entscheiden. Für den Zugang zum Internet gilt nichts wesentlich anderes als für Experimente im Chemieunterricht oder für Expeditionen im Geographieunterricht."

Das vollständige Protokoll der Online-Konferenz, in die sich über 500 CompuServe-Nutzer eingeschaltet hatten, kann aus dem SPIEGEL-Forum geladen werden.

Andere Foren enthalten deutsche, amerikanische, britische und französische Nachrichten, das Wetter, Filmkritiken, usw.

CNNFORUM	CNN-Nachrichten
APO	Associated Press
OLT	On-Line Today
USNEWS	Nachrichten
APFRANCE	französische Nachrichten
DPANEWS	Deutsche Presseagentur
WEATHER	Weltweites Wetter mit aktuellen Satellitenbildern
EBERT	Eberts's Movie Guide: Filmkritiken
EDFORUM	Education

Ein Gateway zum Internet ermöglicht seit einigen Monaten den Zugang zu den USENET-Newsgroups des Internet.

3.2.4 ECHO - European Commission Host Organisation

ECHO ist eine nichtkommerzielle Organisation, die Datenbanken in allen Amtssprachen der Europäischen Union anbietet. Die Aufgaben von Echo sind, über Informationsdienste in der EU zu informieren und Benutzer durch Anleitung, Ausbildung und die Bereitstellung von Dokumentationsmaterial zu unterstützen. Die meisten Echo-Datenbanken sind kostenlos.

[4] Quelle: SPIEGEL-Forum, CompuServe

Sie lassen sich in vier Hauptkategorien einteilen:

- Benutzerhilfe: Datenbanken, Hosts, Produzenten und Informationsvermittler in Europa
- Gemeinschaftsforschung: Gemeinschaftsprogramme für wisser.schaftliche und technische Forschung
- Terminologie: Begriffe und Abkürzungen in allen Gemeinschaftssprachen
- Wirtschaft: Wechselkurse und Ausschreibungen.

Dazu kommen Übungsdatenbanken, zugänglich über Datex-J, Datex-P oder eine Einwahlnummer in Luxemburg.[5]

4 Das Internet

Über **das** Netz ist in den letzten Monaten auch in Tageszeitungen und Nachrichtenmagazinen soviel geschrieben worden, daß wir uns hier auf schulische Nutzungsmöglichkeiten beschränken. In der amerikanischen Zeitschrift *The Computing Teacher* erscheint seit etwa zwei Jahren die Kolumne *Mining the INTERNET*. Hier werden praktische Ratschläge gegeben, mit welcher Software unterrichtsrelevante Informationen von welchen Servern geladen werden können. Speziell auf Schulen zielt eine neue Zeitschrift, *Classroom Connect*, die erfolgreiche Beispiele aus der Schulpraxis vorstellt und praktische Hilfen anbietet. Auch Email-Adressen für elektronische Schulpartnerschaften werden vermittelt.[6]

Neue benutzerfreundliche Software und damit verbunden der Verzicht auf kryptische Befehle machen das Internet auch für Schulen interessant. Ein Teil des Netzes, das World Wide Web (WWW), wird es auch bei uns innerhalb weniger Jahre zu einer unverzichtbaren Ressource für die Arbeit in der Schule machen. Der Hypermedia-Dienst hat enorme Zuwachsraten und ist enorm ausbaufähig. Texte, Bilder, Audio- oder Videodateien sind im WWW über Schlüsselwörter (Hyperlinks) miteinander vernetzt, die bei Bedarf eingesehen oder gehört werden können. Sie werden einfach dazugeladen. Da sie in der Regel auch Hypermedia-Dokumente sind, erfolgen von ihnen aus weitere Verzweigungen. Das Gute für den Anwender ist, daß er nicht wissen muß, auf welchem Rechner sich die aktuelle Datei befindet.

Judi Harris (a. a. O.) berichtet über ein Internet-Projekt der Hillside Elementary School. Schüler der 6. Klasse hatten den Auftrag, zu verschiedenen Themen Material zu sammeln und dieses als WWW-Seiten aufzubereiten. Unter

[5] Kontakt: ECHO Customer Service, B.P. 2373, L-1023 Luxembourg, Tel. 0130-823456

[6] Fax 001-717-393-5752, Email: conect@wentworth.com

ihrer WWW-Homepage[7] kann sich jeder Internet-Nutzer die Ergebnisse ansehen. Eine Gruppe hat ihre mit MOSAIC in HTML (Hypertext Markup Language) über Dinosaurier geschriebenen Seiten mit auf einem Server in Berkeley gespeicherten Informationen verbunden.

Die Schüler haben hiermit in ihrem Projekt die Arbeitsergebnisse auf eine bisher nicht mögliche Weise einem weltweiten Publikum zugänglich gemacht. Auch Lehrer beginnen, über das WWW ihren Schülern Informationen aufzubereiten. Unter ihrer Homepage[8] kann das Material abgerufen werden.

Bleibt zu hoffen, daß der Nachfragedruck aus deutschen Schulen so anwächst, daß Netzwerkanbieter, Kultusbürokratie und Schulträger zu einer Gesamtlösung finden, die es den Schulen ermöglicht, sich anzukoppeln. In den skandinavischen Ländern ist man schon einen großen Schritt weiter und bietet interessierten Schulen zu erschwinglichen Preisen einen vollen Internet-Zugang.

5 Datenbanken auf CD-ROM

Die hohe Speicherkapazität der kleinen Silberscheibe ermöglicht es, ganze Zeitungsjahrgänge, 25-bändige Lexika und Wörterbücher und/oder eine Vielzahl von Bildern, kurzen Audiosequenzen oder Videoclips darauf unterzubringen. Die Materialkosten sind im Vergleich zum Buchdruck extrem gering, und mit der Höhe der Auflage sinken häufig die Preise in für die Schule interessante Regionen. Der Vorteil gegenüber den Online-Diensten ist der Wegfall von Telekommunikationskosten. Der Nachteil, die geringere Aktualität, ist im Schulbetrieb von geringerer Bedeutung.

Eine Vorreiterrolle bei der Einführung von CD-ROM-Applikationen in den Unterricht spielte wiederum England. Eine Secondary School in England verfügte schon vor einigen Jahren über 50 oder mehr Computer, die häufig vernetzt sind. Es lag nahe, die Einsatzmöglichkeiten für CD-ROM-Datenbanken in einem großen vom National Council for Educational Technology (NCET) koordinierten Modellversuch zu evaluieren.[9] Von Dezember 1991 bis Februar 1992 erprobten 30 Grundschulen und 523 Sekundarschulen die neue Technologie. Zur CD-ROM-Ausstattung gehörten u.a. *Grolier's Electronic Encyclopedia*, *World Atlas*, *Library of the Future* (eine Sammlung von mehr als 1000 Werken der Weltliteratur), *The Times* and *The Sunday Times*, *The Guardian*. Bei den Schülern war große Begeisterung festzustellen. Die Schulbibliothekare berichteten, daß Schüler, die sonst nie in den Räumen gesehen wurden, mit den CD-ROM-Datenbanken arbeiteten. Die Rückmeldungen aus

[7] http://hillside.coled.umn.edu/research/papers.html
[8] http://www.imsa.edu/edu/geophysics

den Schulen waren so positiv, daß das Department for Education für die Finanzjahre 1992 und 1993 jeweils 4 Mio. Pfund zur Verfügung stellte, um alle Secondary Schools mit CD-ROM-Laufwerken und Software auszustatten. 1994 folgte ein 4,5 Mio. Pfund Programm, um 2000 Grundschulen mit Multimedia-Computern und CD-ROMs zu versorgen. Eine unabhängige Evaluationsgruppe wählte 20 Titel nach folgenden Kriterien aus:

Kategorien	*Evaluationskriterien*
- Enzyklopädien/Nachschlagewerke	- Übereinstimmung mit dem National Curriculum
- Atlanten und Führer	- Inhalt und Reichweite
- Wörterbücher	- altersbezogene Lesefertigkeiten
- Literarische Werke	- Benutzer-Interface, Präsentationsformen Interaktivität, Möglichkeiten
- Zeitungen	- Angemessenheit des Zusatzmaterials
- Statistische Daten	- Gesamtqualität des Titels
- Bibliographien	
- sonstiges Unterrichtsmaterial	

Ausgewählt wurden u.a. *Anglo Saxons, Busy Town, Creepy Crawlies, Information Finder World Book Encyclopedia, Just Grandma and me, Mammals Multimedia, Microsoft Dinosaurs, Microsoft Encarta, Microsoft Art Gallery, Microsoft Musical Instruments, New Kid on the Block.* Einige diese Titel sind hervorragend für den Einsatz im Englischunterricht oder im bilingualen Fachunterricht geeignet. Deutschsprachige CD-ROM-Titel, die im Unterricht eingesetzt werden könnten, sind noch rar oder zu teuer. Folgende Titel kommen zum Beispiel in Betracht:

Frankfurter Allgemeine Zeitung auf CD-ROM
TAZ auf CD-ROM
Bild der Wissenschaft
verschiedene *Duden*-Wörterbücher
Deutsche Klassiker (erscheint im Frühjahr 1995)

Noch werden in Deutschland erst an ganz wenigen Schulen Online- oder CD-ROM-Datenbanken eingesetzt. Es bleibt viel zu tun, um hier den Vorsprung anderer Länder aufzuholen.

[9] Kontakt: NCET, Milburn Hill Road, Science Park, Coventry CV4 7JJ, Tel 0044-203-416994

6 Perspektiven

Für den Unterricht zeichnet sich der Beginn eines Umbruchs ab.

- Das traditionelle Lehrwerk wird an Bedeutung verlieren. Interaktive Multimedia-Lernsysteme, die den Zugriff auf Texte, Bilder, Ton und Filmsequenzen ermöglichen, elektronische Post, Datenbanken, Computer-Konferenzen, online-Hypermedia-Datenbanken (WWW) werden zunehmend wichtiger und eröffnen neue didaktische und methodische Perspektiven.
- Schüler benutzen "echte" Systeme, wie sie auch in der Arbeitswelt eingesetzt werden. Sie überwinden damit die Isolation des Klassenraumes und heben die Künstlichkeit der Lernsituation auf.
- Internationale Kommunikation und die Teilnahme an internationalen Projekten ermöglichen die zweckgerichtete Anwendung von Fremdsprachen in einem realen Kontext.
- Elektronische Post überwindet politische und kulturelle Grenzen und ermöglicht interkulturelles Lernen. Die Schüler entdecken Gemeinsamkeiten, aber auch kulturelle Unterschiede und lernen, diese zu respektieren.
- Die globalen Herausforderungen Umweltschutz, Klima, Regenwald, Krieg wurden in einer Reihe von Projekten aufgegriffen. Die Schüler erkennen gemeinsame Interessen und die Notwendigkeit gemeinsamen Handelns.
- Der Lernstil ändert sich. Lernen wird zunehmend ein kooperativer Prozeß, sowohl innerhalb der Gruppe / Klasse als auch zwischen elektronisch vernetzten Gruppen. Mehr Selbständigkeit und die Möglichkeit, Entscheidungen zu treffen, führen zu mehr Verantwortungsbewußtsein bei den Schülern.
- Aufgaben und Rolle des Lehrers sind neu zu definieren. Die Rollenvorstellung des lehrbuchgestützten, omnipotenten Wissensvermittlers ist nicht mehr zeitgemäß. Der Lehrer schafft Rahmenbedingungen und wird zum Berater.

Anschrift des Verfassers:

Wolfgang Meyer
Ölmühlenweg 2
D-26506 Norden
Tel. (04931) 21 46; Fax: (04931) 16 80 95
E-Mail: 100345.2672@CompuServe.com

Peter Diepold

Elektronische Informationsdienste für Wissenschaftler am Beispiel des Internet

Was braucht der Wissenschaftler?

Er muß

- die Übersicht zum Stand der Forschung in seinem Gebiet behalten, also insbesondere **über** neuere Arbeiten informiert werden,
- die neueren Arbeiten selbst, insbesondere Aufsätze und Bücher, möglichst schnell erhalten,
- seine eigenen Konzepte, Fragestellungen und Ergebnisse öffentlich machen und
- sie mit Fachkollegen diskutieren.

Das ist schwierig, denn es gibt zwischen 100.000 und 300.000 Zeitschriften auf der Welt; das sind 3 bis 10 Millionen Aufsätze im Jahr. Täglich werden 2.000 Bücher veröffentlicht; davon allein ein Drittel auf Englisch, Deutsch oder Französisch (230.000 jährlich). Täglich produzieren Wissenschaftler weltweit 7000 Arbeiten. Jährlich werden 800.000 Patente angemeldet.[1] Kann er diese Informationsflut noch beherrschen, oder gilt: Er dürstet nach Wissen, aber ertrinkt in Informationen?

Wie können ihm elektronische Informationsdienste seine Aufgabe erleichtern?

Um diese Frage zu beantworten, werde ich im 1. Teil etwas Grundsätzliches zu elektronischer Information und Kommunikation ausführen, im 2. Teil die im Internet verfügbaren Informations- und Kommunikationsdienste für den Wissenschaftler beschreiben, im 3. Teil die Nutzung an Beispielen aus dem eigenen Arbeitsbereich darstellen und im 4. Teil einige Anmerkungen zu den Möglichkeiten des elektronischen Publizierens machen.

---

[1] Vgl. E.-G. vom Kolke: Online-Datenbanken, München 1994, 1; B . Kurshan: Creating the Global Classroom for the 21st Century. In: Eric Persson: Net Power. Resource Guide to Online Computer Services. Lancaster PA, 1993, S. iii; L. A. Tsantis: Collaborative Networking. An Educator Looks at the New Technologies. Ebd., S. viii

1 Elektronische Information und Kommunikation

1.1 Der informationelle "Quantensprung" des 20. Jahrhunderts

In der Geschichte menschlicher Informationsverarbeitung hat es mehrere "Quantensprünge" gegeben, die Information und Kommunikation auf eine qualitativ neue Stufe gehoben haben.

- Der erste informationelle Quantensprung ist die Entwicklung der Schrift, genauer: die Erfindung des Alphabets im 2. vorchristlichen Jahrtausend im semitischen Sprachraum. Er ermöglichte es, sprachliche Informationen durch nur 22 Zeichen kodierbar zu machen.
- Der zweite ist Gutenbérgs Druckpresse zu Beginn der europäischen Neuzeit. Er ermöglichte die mechanische Vervielfältigung von Wissen.
- Der dritte ist die Informations- und Kommunikationstechnik des 20. Jahrhunderts.

Was **elektronische** Informationsverarbeitung bedeutet, läßt sich an einem Vergleich zu traditioneller Informationsverarbeitung nachvollziehen.

Verarbeitungsgeschwindigkeit. Ein mit 40 MHz "getakteter" Tischcomputer kann in einer Sekunde etwa 10 Millionen Befehle ausführen. Nehmen wir an, ein Mensch schaffte zwei Befehle in der Sekunde, dann bräuchte er dafür 5 Mio Sekunden, das sind 83.000 Minuten oder 1.400 Stunden oder 36 Arbeitswochen. Wozu der Computer eine Sekunde benötigt, braucht ein Mensch also ein dreiviertel Jahr.

Um den Inhalt einer CD-ROM mit Volltexten zu indizieren, wären 833 Mannjahre nötig; der Computer braucht dazu 20 Minuten.

Speichermöglichkeit. Eine CD, ein inzwischen weitverbreitetes Speichermedium für Bibliographien, Volltexte, Grafiken und Bilddateien, faßt 660 Megabyte an Informationen. Das entspricht dem Inhalt von 45 lfd. Meter Stellfläche, nämlich 1.500 Büchern. Im Vergleich zum traditionellen Speichermedium, dem Buch, benötigt die CD an Platzbedarf, Herstellungskosten, Materialaufwand weniger als ein Tausendstel. In Verbindung mit der Verarbeitungsgeschwindigkeit von Computern ist sie dem Papiermedium millionenfach überlegen (vgl. Abb. 1: "Technologievergleich CD-ROM <-> Buch").

Während in den vergangenen Jahrhunderten das menschliche Wissen in Büchern festgehalten und in Bibliotheken gesammelt wurde, wird heute das Welt-Wissen zusammengefaßt in derzeit ca. 6.000 großen Datenbanken, ist hervorragend strukturiert und indiziert, so daß beliebig kombinierbare Ab-

Technologievergleich CD-ROM <-> Buch

	1 CD-ROM	*1.500 Bücher*	*Faktor*
Materialbedarf	16 g Polykarbonat	30 Bäume	
Herstellungszeit	wenige Sekunden	20 Stunden	1 : 10.000
Platzbedarf	1 cm	44 m	1 : 4.400
Gewicht incl. Verpackung	110 g	500 kg	1 : 3.000
Preis	DM 50 - 400	DM 60.000	1 : 1.000
Versandkosten	DM 2,50	DM 600	1 : 240
mittlere Zugriffszeit	200 ms	mehrere Minuten	1 : 1.000
einmal durchlesen	20 Minuten	10 Jahre	1 : 260.000
vollständige Indizierung	20 Minuten	800 Mannjahre	1 : 21 Mio.
Kosten der Indizierung	DM 0,02	DM 40 Mio	1 : 2 Mrd.

fragen in Sekundenschnelle die Antworten liefern. Viele Datenbanken enthalten nicht nur die Information **über** Quellen (wie z. B. Literaturdatenbanken oder die elektronischen Kataloge der öffentlichen Bibliotheken), sondern die Quelle selbst als sog. Volltext-Datenbanken. Von ihnen kann man Auszüge als elektronische Kopie auf Diskette, als Fax oder als "Download" auf den eigenen Rechner erhalten.[2]

Solche für Erziehungswissenschaftler wichtige Datenbanken sind beispielsweise:

- die deutsche Nationalbibliographie,
- Dissertation Abstracts,
- Forschungsinformationssystem Sozialwissenschaft (FORIS),
- Mathematik Didaktik (MATHDI),
- aktuelle Umweltdaten der Umweltlandesämter,
- Datenbanken der Presseagenturen (ap, Reuters, dpa),
- Texte der überregionalen Zeitungen (FAZ, SZ),
- System für Information über Graue Literatur in Europa (SIGLE),
- Sozialwissenschaftliches Literaturinformationssystem (SOLIS),
- Genios-Wirtschaftsdatenbanken,
- die OPACS ("online public access catalogs") der wissenschaftlichen Bibliotheken).

Datenverarbeitung. Die fast nicht begreifbare Schnelligkeit des PC macht es möglich, in riesigen Datenmengen komplexe Suchaufträge sofort - "in Echtzeit"- durchzuführen. In einem Augenblick - einer Viertelsekunde - hat der PC eine Million Befehle abgearbeitet, Hunderte von Seiten Text gelesen, Textteile miteinander kombiniert, verglichen, bearbeitet, Indices in jeder gewünschten Reihenfolge mittels Boolscher Operatoren kombiniert und die Ergebnisse auf dem Bildschirm ausgegeben, damit man sie auf Papier ausdrucken oder auf einer Diskette abspeichern kann.

Vernetzung. Ist der Computer mit anderen Computern verbunden, kann er mit ihnen in Sekundenschnelle Kontakt aufnehmen, deren Speicher und Verarbeitungsfähigkeit nutzen und Daten übertragen, Ergebnisse übermitteln. Verbunden mit den in Tausenden von Datenbanken liegenden Informationen, wird der Tischcomputer Teil eines weltumspannenden elektronischen "Gehirns" mit millionenfach gesteigerter Gedächtnis- und Denkleistung.

---

[2] Diese Informationen sind als Dienstleistung natürlich nicht kostenlos, doch gibt es für den Bildungsbereich in der Regel Abschläge, nicht selten bis 90 % des Normalpreises (z. B. beim FIZ/STN Karlsruhe, bei dem großen amerikanischen Datenbankanbieter DIALOG (400 Datenbanken, 300 Mio Dokumente, 1.100 Volltext-Zeitschriften: "Dialog Classmate"). Online (direkt) erreichbar sind sie für den Computer des Lehrers technisch über Modem und Telefonanschluß, organistorisch über Dienste wie Datex-J (Datex für Jedermann, das alte BTX), über kommerzielle Anbieter wie CompuServe oder über das Internet.

Hier liegen die besonderen Möglichkeiten elektronischer Informationsverarbeitung. Sie sollen im folgenden am Beispiel des im Bereich der Wissenschaft weitverbreiteten INTERNET erläutert werden.

1.2 Entstehung und Struktur des Internet[3]

Die Geburtsstunde des Internet - so ein weiterer Mythos mit einem wahren Kern - war Ende der 60er Jahre, als das amerikanische Verteidigungsministerium einen Auftrag erteilte, Hochleistungscomputer so zu verbinden, daß diese Verbindung trotz eines Atomangriffs weiter funktionieren könnte. Die Lösung war so genial wie einfach:

Network. Es wurde ein weitverzweigtes, "vermaschtes" **Netz** von möglichen physischen Kommunikationsverbindungen genutzt: Daten- und Telefonleitungen, Funkverkehr, Satellitenverbindungen. Viele Wege führen von einem Computer zum anderen. Sollte einer unterbrochen werden, wird eine Umleitung gewählt.

Protocols. Es wurden verbindliche Vereinbarungen über die Kommunikation zwischen Computern getroffen. Diese "Protokolle" abstrahieren von der zugrundeliegenden Hard- und Software. Das bedeutet: Im Gegensatz zu proprietären Netzsystemen wie AppleTalk, Novell, Banyan Vines oder Microsoft können PCs, Mac oder Unix-Rechner miteinander kommunizieren. Die Protokollfamilie des TCP/IP ist quasi die *lingua franca* des Internet geworden.

Packet switching. Der Datenstrom wird in kleine Pakete aufgeteilt, die mit der Adresse von Absender und Empfänger versehen, unabhängig voneinander auf den Weg geschickt werden. Wird eines beschädigt oder geht es verloren, so wird es automatisch erneut übertragen; es ist nicht nötig, daß die *gesamte* Kommunikation wiederholt wird. Darüber hinaus ermöglicht der Paketdienst, daß Leitungen gleichzeitig von mehreren Benutzern belegt werden können.

Name Service. Jeder Rechner hat einen Namen und (ähnlich einer Telefonnummer) eine weltweit eindeutige vierteilige numerische Bezeichnung. So hat z. B. unser Server der Abteilung Pädagogik und Informatik der Humboldt-Universität den Namen `"bonsai.educat.hu-berlin.de"` bzw. die Internet-Nummer `"141.20.130.40"`

Systemtheoretisch verblüffend ist die Analogie zu natürlichen Systemen, wie sie in der Biologie vorkommen, z. B. verteilte, ganzheitliche Informationsverarbeitung und -speicherung in den neuronalen Netzen des Gehirns, streng chaotische Struktur, dezentrale Steuerung durch das Zusammenwirken aller

[3] Zur Geschichte des Internet vgl. Henry E. Hardy, The History of the Net. umcc.umich.edu /pub/seraphim/doc/nethist8.txt.

Teile, große Flexibilität, Fähigkeit, sich neuen Erfordernissen schnell und ohne Störung anzupassen, passive und aktive Mitwirkung aller daran beteiligter Komponenten.

Bereits 1972 wurde dieses militärische ARPANET[4] freigegeben und entwikkelte sich zu einem Forschungsnetz, dem Internet. Zunächst nur von wenigen Insidern in Universitäten und Forschungsinstituten genutzt, hat es sich in den letzten Jahren bei exponentiellem Wachstum zu einem weltumspannenden Kommunikationsnetz entwickelt. Seit wenigen Jahren ist dieses Netz auch deutschen Wissenschaftlern zugänglich, nachdem der Verein Deutsches Forschungsnetz zunächst lange die Protokolle der OSI[5] favorisiert hatte.

In USA wurde die Entwicklung des Internet durch eine offensive Bildungs- und Forschungspolitik gestützt. So finanziert die National Science Foundation den Anschluß einer Universität an das Internet nur, wenn die Hochschule auch alle Arbeitsplätze und studentischen Computer-Räume lokal vernetzt und mit dem Internet verbindet. Die Regierung Clinton hat Ende 1993 ein Fünf-Punkte-Programm vorgelegt, das u. a. den freien Zugang zum Netz für alle Bürger unter dem Postulat von freiem Wettbewerb und Flexibilität vorsieht.

Derzeit sind im Internet schätzungsweise 30 Millionen Computer in allen Kontinenten und in mehr als 50 Ländern weltweit verbunden. Andere Netze, wie EARN, BITNET, CompuServe, Fidonet oder NewsNet, haben sog. "Gateways", über die sie mit dem Internet in Verbindung treten und einige Dienste nutzen. "Peer-to-peer-communication" ist möglich, "jeder Benutzer des Netzes", so Vizepräsident Al Gore, "kann sowohl Konsument als auch Produzent von Information sein".[6]

Für den Zugang zum Internet mieten die Hochschulen i. d. R. eine Standleitung vom Verein Deutsches Forschungsnetz. Die Gebühren dafür sind in Deutschland wegen des Monopols der Deutschen Telekom außerordentlich hoch; sie beginnen bei DM 25.000 und gehen bis über DM 400.000 im Jahr. Über die Universitätsrechenzentren erhalten dann die Mitglieder der Universität - Dozenten, Mitarbeiter, Studenten - auf Antrag kostenlosen Zugang zum Internet.

Für private Nutzer[7] vermitteln zwei Vereine ihren Mitgliedern den Internet-Zugang zu moderaten Preisen, die sich nach dem Umfang an Diensten und

---

4 Advanced Research Projects Agency Net

5 Open Systems Interconnections

6 EFFector Online, vol. 6, no. 8, 12/28/1993

7 IBM, Microsoft und Apple bieten in USA inzwischen beim Kauf ihres jeweiligen Betriebssystems Internet-Zugänge als Gutschein mit 30 Tagen Probe über eigene Zugangspunkte (points of presence) an.

dem Volumen orientieren.[8] Neuerdings drängen auch kommerzielle Anbieter in das Geschäft und verändern damit den Charakter des Internet als eines Wissenschaftsnetzes.

Technisch kann der Internet-Zugang auf zwei Weisen verwirklicht werden: (1) Der Arbeitsplatz-PC ist fest per Standleitung eines lokalen Netzes mit dem Rechenzentrum verbunden; oder (2) der Computer ist an einen Modem angeschlossen, mit dessen Hilfe man sich über eine Telefonverbindung - vom Arbeitsplatz, von zu Hause, von unterwegs - in das Rechenzentrum einwählt.

Mit dem Internet kann nun der Wissenschaftler auf die elektronischen Speicher von Millionen Computern, gefüllt mit dem Wissen der Welt, zugreifen, die (oben beschriebene) immense Rechnerleistung nutzen und mit anderen Wissenschaftlern in Kontakt treten. Welcher Dienste er sich dabei bedient, werde ich im folgenden zweiten Teil erläutern.

2 Informationsdienste für den Wissenschaftler

2.1 TELNET

Mit dem Befehl "Telnet", gefolgt von der vierteiligen Rechnernummer, nehme ich Kontakt zu einem externen Rechner auf. Über das Internet sind die **Bibliotheken** der Welt mit ihren bibliographischen Hilfsmitteln erreichbar, Katalogen, Bibliographien, CD-ROMS, sei es die nächste Universitätsbibliothek oder die Library of Congress, der Welt größte Bibliothek. Der Zugang ist öffentlich und frei.

So können Wissenschaftler folgende Dienste benutzen:

- die elektronischen Kataloge (OPAC)[9], über die nicht nur die Bestände der öffentlichen Bibliotheken abgefragt, sondern auch Buchbestellungen aufgegeben werden könnten,
- das deutsche Verzeichnis lieferbarer Bücher oder das englische books in print,
- Datenbanken wie die Deutsche Bibliothek oder andere Nationalbibliographien,
- bibliographische Datenbanken wie SOLIS[10] oder ERIC[11], die über die Literaturnennung hinaus Schlagworte und Referate (Abstracts) enthalten.

---

[8] Es sind dies Individual Network e.V. (IN), Scheideweg 65, 26121 Oldenburg sowie subNetz e.V., Gerwigstr. 5, 76131 Karlsruhe. Zu Einzelheiten vgl. V. Heinau; H. Schlichting: "Private Zugänge zum Internet", in ct (1994), H. 7, S. 245-248.

[9] Online Public Access Catalogue

[10] Sozialwissenschaftliches Literatursuchsystem

Mit Hilfe von Telnet haben wir im letzten Jahr an meinem Lehrstuhl inzwischen eine fast 7.000 Nachweise umfassende Bibliographie zum Themengebiet "Pädagogik und Informatik" aufgebaut: Es wurde in verschiedenen Datenbanken recherchiert, die Ergebnisse elektronisch in den eigenen Computer gebracht und von dort mit einem Umsetzungsprogramm[12] in unser eigenes Format überführt, mit dem wir unter LIDOS arbeiten. Das hat uns Hunderte von Arbeitsstunden für die Datenaufnahme von Literaturnachweisen erspart. Aus dieser eigenen Datenbank haben wir dann auch gleich die Bestellungen für die neu aufzubauende Bibliothek der Abteilung ausgedruckt.[13]

2.2 FTP

Bibliotheksdatenbanken sind nur ein Beispiel für die **Datenbanken**, die das Internet bereithält. Die Zahl der erreichbaren Datenbestände geht in die Legion - seien es die Satelliten-Wetterkarten, die Datenbanken der Genome oder Tausende **Programme** für DOS und Windows, für UNIX oder OS/2, für Macintosh, Atari oder Amiga. Es gibt kaum ein Problem, für das nicht irgendwo eine Lösung existiert, die man - im Wissenschaftsbetrieb des Internet die Regel - kostenlos erhalten kann.

Ähnliches gilt für Volltexte. Das Projekt GUTENBERG der University of Illinois hat sich das ehrgeizige Ziel gesetzt, 10.000 Titel der Weltliteratur elektronisch aufzunehmen und im Internet bereitzuhalten. Das Projekt DANTE an der Dartmouth University hat den italienischen Text der Göttlichen Komödie und wissenschaftlichen Apparat der Sekundärliteratur aus mehreren Jahrhunderten aufgenommen.

Von der mit 16 Mio Büchern größten Bibliothek der Welt, der Library of Congress, war im SPIEGEL[14] kürzlich zu lesen, daß ihre Bestände nach und nach als Volltexte verfügbar gemacht werden sollen; bis zum Jahr 2000 sollen 5 Mio Dokumentationseinheiten elektronisch als Volltexte abrufbar sein.

---

[11] Educational Resources Information Center

[12] Hilfsprogramm "Download" des von der Doris Land Software vertriebenen Literaturverwaltungsprogramms LIDOS

[13] Unser Vorgehen - von der elektronischen Recherche bis zur Überführung der Ergebnisse in das eigene Literaturverwaltungsprogramm - haben wir auf einem Workshop der GIB im Mai 1994 in Berlin vorgestellt.

[14] *DER SPIEGEL* 43/1994, S. 240-243

2.3 E-Mail

Electronic-Mail. Über das Netz ist die schnelle Kommunikation mit anderen Wissenschaftlern möglich, wo immer sie sich auf der Welt befinden. Das ist hilfreich, wenn man z. B. längere Texte gemeinsam verfassen oder herausgeben möchte. Anders als beim Fax brauchen die empfangenen Texte nicht noch einmal eingetippt werden; Korrekturen können am Computer vorgenommen und der Text kann sogleich wieder zurückgeschickt werden - ein großer Vorteil, wenn Termine anstehen und man Druckfehler vermeiden möchte.

Die Entwicklung geht jetzt in Richtung auf Multimedia-Mail, die nicht nur Texte, sondern auch Tabellen, Zeichnungen, Bilder, Videos oder Tondateien übertragen kann.

2.4 Diskussionsforen und E-Mail-Listen

Während E-Mail als private Post zwischen Einzelnen verstanden werden kann, gibt es E-Mail-Listen und Diskussionsforen, in denen die Beiträge der Absender öffentlich gemacht sind. Internet bietet derzeit mehr als 6000 thematische Diskussionsforen und Hunderte von E-Mail-Listen an, die Wissenschaftler kostenlos abonnieren können. Zu den für Erziehungswissenschaftler interessanten gehören beispielsweise:

- European Association for Research in Learning and Instruction
- Adult Education Network
- Alternative Approaches to Learning
- American Educational Research Association
- Curriculum Development Issues
- Educational Policy Analysis Archives
- Education Research List
- The Science of Education
- Swedish Initiative for Research in Education
- Society of Teaching and Learning in Higher Education
- Education for talented and gifted students
- Education Research List

Dabei spielen Fragen an die Kollegen oder Hilferufe eine nicht unbeträchtliche Rolle: Eine neue Problemstellung, eine Sachfrage, Aufforderung zur Beteiligung an einem gemeinsamen Projekt, Stellenangebote und -suche werden in der Regel schnell beantwortet, spricht man in einem Forum doch die gesammelte Expertise von Kollegen in Europa und Übersee an.

2.5 Elektronische Zeitschriften

Die meisten Diskussionsforen sind frei, manche werden moderiert. Hier deutet sich schon der Übergang zu den neuen **elektronischen Zeitschriften** an. Mit der internationalen Seriennummer (ISSN) versehen und von Fachkollegen redigiert, unterscheiden sie sich von den üblichen wissenschaftlichen Zeitschriften in ihrer Aktualität und natürlich in ihrem Verbreitungsgrad. Das Abonnement der meisten von ihnen ist kostenlos, wie üblicherweise auch das aller Diskussionsforen.

Hier einige Beispiele, die für Pädagogen interessant sind:

- PSYCHOLOQUY. Refereed Electronic Journal of Peer Discussion
- CATALYST. The Community Services
- JTE-L. Journal of Technology Education
- INTERACT. European Platform for Interactive Learning.

2.6 Informationsdienste des Internet

Für den Wissenschaftler stellt sich anfangs das Problem, wie er sich in dieser Informationsflut zurechtfinden soll. Es gibt inzwischen praxisorientierte Einführungen, die ihm die ersten Schritte zeigen. Darüber hinaus stellt das Internet Informationen über Informationen zur Verfügung.

ARCHIE. Bevor man Informationen abrufen kann, muß man wissen, wo sie liegen. ARCHIE, ein weltweiter Index, greift auf die Inhaltsverzeichnisse aller öffentlicher Server im Internet zurück und verwaltet Informationen über 2,5 Millionen öffentlich verfügbarer Dateien, ihre Namen, Größe, letzte Bearbeitung sowie Fundort. Diese Informationen werden gleichzeitig auf zwei Dutzend Servern auf der Welt vorgehalten; sie sind so etwas wie ein großer OPAC aller verfügbaren Dateien.[15]

WAIS. Während ARCHIE das Verzeichnis der Namen aller verfügbarer Dateien verwaltet, ist WAIS ("wide area information service") ein Volltext-Recherche-Dienst, der Text-Dateien indexiert und es ermöglicht, das Internet nach inhaltlichen Stichworten zu durchsuchen.

GOPHER. Hinter diesem Dienst verbirgt sich ein leistungsfähiges Menüsystem, mit dessen Hilfe sich auch Computer-Ungeübte im Internet zurechtfinden, z. B. Recherchen in Bibliotheken durchführen, ohne daß sie sich mit Telnet, FTP und der Eingabe von Adressen abmühen müssen.

[15] Der Name des deutschen Archie ist: archie.th-darmstadt.de.

WORLD WIDE WEB. Noch einfacher geht es mit World Wide Web, einem Hypertextsystem im Internet. Wissenschaftliche Texte verweisen in Fußnoten mit "s.o." oder "s.u." auf andere Textstellen sowie auf zitierte, benutzte oder weiterführende Literatur. Solche Verweise können natürlich auch in elektronischen Texten stehen. Warum dann nicht gleich auch den Text selbst liefern? Kein Problem, wenn ein entsprechender Verweis auf die Fundstelle, ein "Link", gesetzt ist. Die über einer Menge von Daten liegende Struktur von Links nennt man Hypertext, und wenn nicht nur Texte, sondern auch Grafik, Bilder, Tondateien, Videosequenzen eingebunden sind, sprechen wir von Hypermedia. Das World Wide Web ist ein solches hypermediales Informationssystem, das sich nicht auf den eigenen Computer beschränkt, sondern, wie der Name sagt, alle Computer im Internet einschließt, die auf der Basis von HTML ("hypertext markup language") Daten zur Verfügung stellen.

3 Electronic Publishing

Das WWW markiert den Übergang zur **aktiven** Nutzung des Internets. Bisher habe ich - eher aus dem Blickwinkel des **passiven** Nutzers von Informationsdiensten - über die Nutzung externer Systeme gesprochen, den Zugriff auf Bibliographien, auf Volltexte, auf elektronische Zeitschriften und elektronische Bücher. Aber alle diese Dienste werden doch von anderen, Wissenschaftlern, Bibliothekaren, Dokumentaren, im Rahmen ihrer Tätigkeit in das System **aktiv** eingebracht. In den Vereinigten Staaten ist es üblich, daß Arbeiten, die aus öffentlichen Mitteln finanziert wurden, auch der Öffentlichkeit kostenlos zur Verfügung gestellt werden. Dies korrespondiert mit der Überzeugung, Information sei keine Ware, sondern ein Bürgerrecht. Was für Bücher und Zeitschriften gilt - freier, kostenloser Zugang für alle in Bibliotheken - soll auch für die elektronischen Informationsdienste gelten.

Dies erklärt die immense Masse an Daten, Programmen, Texten im Internet, in einer Qualität, die, besonders in der UNIX-Welt, kommerzielle Produkte nicht selten übertrifft.

In Deutschland nimmt die Zahl der Wissenschaftler zu, die sich dieser Tradition verpflichtet fühlen. Und in dem Maße, in dem Erziehungswissenschaftler sich am Internet beteiligen, steigt die Attraktivität der Möglichkeit, elektronisch zu publizieren. Wissenschaftler beginnen, ihre Schriften - Zeitschriftenaufsätze oder auch ganze Bücher - über FTP zu veröffentlichen. Da die meisten Texte heutzutage ohnehin auf dem Computer geschrieben und erst hinterher auf Papier gebracht werden, bietet es sich an, diese Texte auch ONLINE auf dem Server des Universitäts-Rechenzentrums vorzuhalten, in welcher Form auch immer: als reiner ASCII Text, ohne irgendwelche Formatierungen, im Format eines spezifischen Textprogramms wie WINWORD

oder WordPerfect, in einem von verschiedenen Textprogrammen lesbaren Format wie RTF (rich text format) oder TECH in der Unix-Welt, oder schließlich als Druck-Datei im Postskript-Format, die dann einfach mit einem PRINT-Befehl auf dem eigenen Drucker ausgegeben werden kann.

Ich möchte dies an einem Beispiel, wieder aus meinem Institut, exemplifizieren. Wir betreiben seit einigen Monaten aktiv unter Unix einen eigenen World Wide Web Server - es ist ein einfacher 486er PC, wie man ihn auch an Schulen findet. Wenn man von irgendwo in der Welt unter dem Programm MOSAIC die Adresse

http://www.educat.hu-berlin.de

eingibt, meldet sich unser Server mit der in Abbildung 2 wiedergegebenen Bildschirmseite.

Auf dem Bildschirm sind eine Reihe von Begriffen blau unterlegt: Es sind Verweise ("*links*") auf weiterführende Bildschirmseiten, ähnlich wie die Fußnoten in einem Buch. Klickt man sie mit der Maus an, wird ein weiterer Text angerufen.

Unter dem Stichwort "Angebote der Abteilung" (Abb. 3) finden sich weitere Verweise, so auch auf "Publikationen und Programme". Klickt man sie an, werden die derzeit verfügbaren Materialien aufgeführt (Abb. 4). Darunter befindet sich die Expertise für den BMBW zum Thema "Literaturdokumentation Berufliche Bildung" ebenso wie drei Planspiele. Mit einem weiteren Klick kann man sich darüber näher informieren (Abb. 5) und diese Computerprogramme - wie auch den Text der umfänglichen Expertise - in wenigen Sekunden über das Internet auf sein eigenes Gerät kopieren.

Wenn das Publizieren im Internet so einfach ist, lohnt es sich dann überhaupt noch, Bücher zu machen? Mit wissenschaftlichen Büchern ist in der Regel kein Geld zu verdienen. Bei üblicher Kalkulation beträgt der Verkaufspreis eines Buches etwa das Sechsfache der Satz- und Druckkosten, und Auflagen über 1.000 sind im Wissenschaftsbereich eher die Ausnahme. Da ist ein 400-seitiges Buch fast nicht mehr erschwinglich. Die in den letzten Jahren gestiegenen Kosten wiederum verringern den Absatz und damit die Verbreitung des Werkes. Bei den knappen Etats der Bibliotheken kann man noch nicht einmal mehr erwarten, daß die wissenschaftliche Literatur auch angeschafft wird, die verfügbar ist. Eine Diskette dagegen faßt den Inhalt von 3 Büchern, und das für 1,00 DM Hardware- und 2,00 DM Portokosten.

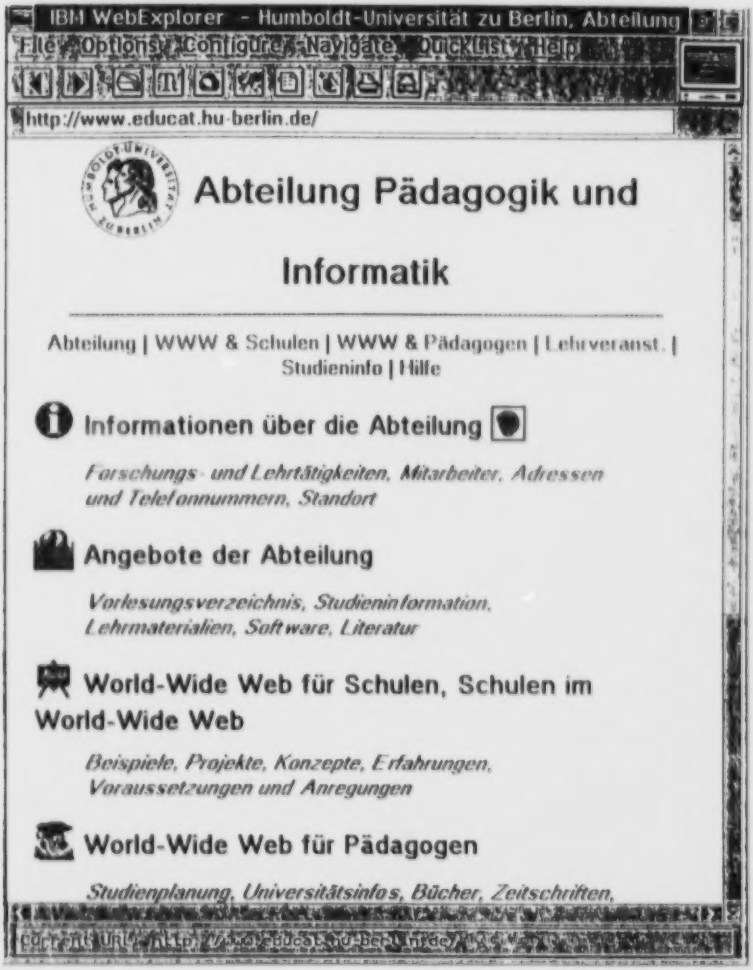


Abb. 2: "Homepage" des Berliner Servers

BEST COPY AVAILABLE

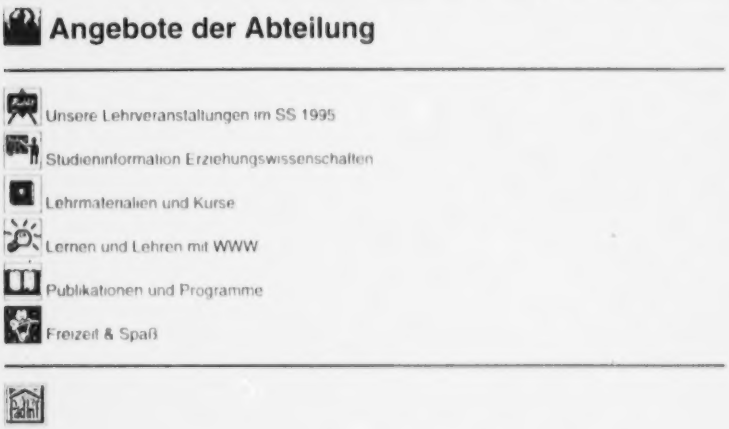


Abb. 3: angewähltes Link: Angebote der Abteilung

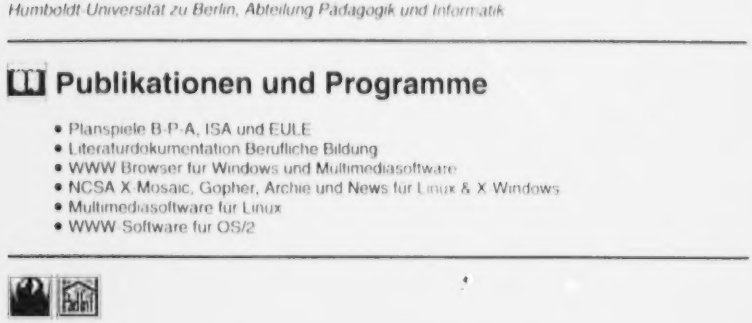


Abb. 4: angewähltes Link: Publikationen und Programme

Planspiele

Im Rahmen des Modellversuchs WOKI wurden drei Planspiele für MS-DOS entwickelt:

- B-P-A modelliert Zusammenhänge zwischen Beschaffung, Produktion und Absatz eines computerproduzierenden kleinen Unternehmens,
- EULE simuliert bereichsspezifisch Einkauf- und Logistik- Entscheidungen eines Reifenproduzenten,
- ISA simuliert den Bereich "Industrie-Simulation Absatz".

Die Planspiele wurden für die Ausbildung von Industriekaufleuten konzipiert, B-P-A ist aber auch für den Einsatz in anderen Ausbildungsgängen bzw. in der Schule einsetzbar; dieses Planspiel wurde ins Englische und Russische übersetzt.

Die drei Planspiele sind Freeware und können bei Wahrung des Copyright (BIBB, Berlin) kopiert und für nichtkommerzielle Zwecke eingesetzt werden.

Falls Ihr WWW-Browser dafür eingerichtet ist, können Sie die Planspiele durch einfaches Anklicken per FTP auf Ihren lokalen Rechner transferieren.

Abb. 5: angewähltes Link: verfügbare Planspiele

Die elektronische Veröffentlichung im Internet ist sogar fast kostenlos (ohne Berücksichtigung der ohnehin anfallenden fixen jährlichen Kosten für das Internet) für den Autor und seine Leser: Die Herstellungskosten beschränken sich auf die ohnehin meist im Institut vorgenommenen Texte und ihre Formatierung mit einem leistungsfähigen Text- oder Desktop-Publishing-Programm. Speicherplatz auf einem Server ist zu finanzieren - pro Buch ist das nicht mehr als maximal ein halbes Megabyte. Bei derzeitigen Preisen von etwa DM 1.- pro Megabyte ist das - über die Lebenszeit einer Platte verteilt - ein Pfennigbetrag.

Darüber hinaus braucht der Autor nicht mehrere Monate auf die Veröffentlichung seines Buches zu warten, Druckfehler können jederzeit korrigiert, inhaltlich notwendige Änderungen sofort vorgenommen werden, die Sache ist aktuell, solange sie gepflegt wird.

Einmal im Internet, sorgen WAIS und ARCHIE für millionenfache Werbung, eine E-Mail an interessierte Kollegen, eine Diskussion der Ergebnisse im entsprechenden Diskussionsforum schaffen eine Verbreitung, die sich ein Buchautor in seinen kühnsten Träumen nicht erhofft und ein Verlag nie finanzieren würde.

Ich erwarte, daß sich in wenigen Jahren die wissenschaftliche Diskussion, die sich bislang in wissenschaftlichen Zeitschriften, in Kongreßberichten und in Monographien niederschlug, zu einem großen Teil elektronisch abspielen wird. Die kommerzielle Verbreitung wissenschaftlicher Ergebnisse auf Papier

wird stark zurückgehen. Dort, wo sich Autoren der Dienstleistung eines Verlages bedienen möchten, wird dies möglich sein: Auch elektronische Bücher lassen sich verlegen und vertreiben, in diesem Falle wird der elektronische Zugang zu ihnen im Internet für den Nutzer nicht mehr frei sein, sondern über das Passwort in Koppelung mit einer Kreditkarte erfolgen. Aber ich denke und hoffe, daß sich die amerikanische Einstellung, wissenschaftliche Informationen frei anzubieten, auch bei uns durchsetzen wird. Das Internet bietet die Struktur dafür, und die Entwicklung der elektronischen Informationsverarbeitung bewirkt eine solch immense Reduzierung an Kosten bei gleichzeitiger Zunahme an Leistung, daß alles andere angesichts knapper öffentlicher Mittel eine Verschwendung von Steuergeldern darstellen würde.

Wohin geht die Reise? Angesichts der Wachstumsraten von 10 % im Monat klingt eine Prognose aus dem Jahr 1992 nicht unwahrscheinlich, nach der im Jahre 2000 das Internet aus ca. 100 Mio Hostrechnern, 3 Mio Netzen und 1 Mrd. Benutzern bestehen werde.

Wir dürfen gespannt sein - auf die weiteren Entwicklungen, die immensen Probleme und auf die Lösungen. Ich möchte am Schluß nur noch stichwortartig nennen:

- Prognosen für das Internet 2000, die von 100 Mio Hostrechnern, 3 Mio Netzen und 1 Mrd Benutzern sprechen,[16]
- technische Entwicklungen wie Packet Radio-basierte Übertragungssysteme (Übertragung zu Amateurfunk-Konditionen: äußerst preiswert) und Satellitenkommunikation,
- drahtlose Daten-Kommunikation von unterwegs,
- der Niedergang des Telekom-Monopols in Deutschland,
- wesentlich vereinfachte Benutzung elektronischer Dienste: mächtigere Suchwerkzeuge, sog. Knowbots (KNOWledge roBOTS),[17]
- die Konsequenzen für die Bibliotheken und für die wissenschaftliche Dokumentation (z. B. hinsichtlich SUBITO (Bund-Länder-Initiative zur Beschleunigung der Literatur- und Informationsdienste)),
- die Realität von "virtuellen Bibliotheken",
- Probleme des Urheberrechts,
- Gefahren der Verfälschung und der Datensicherheit,
- Probleme des Mißbrauchs von Berechtigungen,
- der Abbau von Arbeitsplätzen im traditionellen Bibliotheks- und IuD-Bereich und die Notwendigkeit neuer Qualifikationen.

Wir haben wohl genug Stoff zur Diskussion. Darauf freue ich mich und danke Ihnen fürs geduldige Zuhören.

---

[16] zit. bei K. Sarnow/O. de Presno: Online-World. Globale Kommunikation. Hannover (Heise) 1994, S. 221

[17] Sarnow/de Presno, a.a.O., S. 226

Literatur zum Einstieg:

Bayer, Rudolf (1994): Plädoyer für eine Nationale Informations-Infrastruktur. In: *Informatik Spektrum*, Bd. 17, S. 302-308

Dietrich, Carsten; Dapper, Thomas (1995): Internet. Das größte Informationsnetz der Welt. Würzburg (Vogel Computer Presse) = Chip Special Anwenderpraxis

Kalin, Sally W.; Tennant, Roy (1991): Beyond OPACs. The Wealth of Information Resources on the Internet. In: *Database*, Bd. 14, H. 4, S. 28-33

Krol, Ed (1994): Die Welt des Internet. Handbuch und Übersicht. (O'Reilly & Associates)

Lammarsch, Joachim; Steenweg, Helge (1994): Internet & Co. Elektronische Fachkommunikation auf akademischen Netzen. Bonn u. a. (Addison-Wesley)

Maier, Gunther; Wildberger, Andreas (1994): In 8 Sekunden um die Welt. Kommunikation über das Internet. 3., überarb. Aufl., Bonn; Paris (Addison-Wesley)

Anschrift des Verfassers:

Prof. Dr. Peter Diepold
Humboldt-Universität zu Berlin
Abt. Pädagogik und Informatik
Geschwister-Scholl-Str. 6
D-10099 Berlin
Tel: (030) 2093-3178; Fax: -3198
E-Mail: PDiepold@educat.hu-berlin.de

Robert Funk

Zusammenfassung der Diskussion zu den Vorträgen von Meyer und Diepold

In Hinblick auf die bestehende und sich noch erweiternde Informationsflut ergeben sich für eine Nation, die vom Wissenstransfer abhängig ist und immer abhängiger wird, Probleme mit unüberschaubaren Folgen. Diese Probleme lassen sich mindern, indem man das Informationsbewußtsein insbesondere bei Lehrern und Schülern verstärkt. Um sach- und zeitgerecht auszubilden, ist es unerläßlich geworden, die elektronische Datenverarbeitung in die Ausbildung einzubeziehen. Die Notwendigkeit hierzu ist bisher erkannt, aber durch verschiedene Schwierigkeiten ist die Einbeziehung nur unvollständig realisiert. Das trifft auch auf die Ausbildung zur Arbeit mit elektronischen Literatur- und Informationsdatenbanken zu. Dieser Mangel, der auch schon bei den gedruckten Auskunfts- und Informationsmitteln bestand, hat sich in unserer Diskussion gezeigt.

Folgende Beiträge erscheinen wichtig:

- Das für Schulen geeignete elektronische Informationsmaterial ist nicht sehr umfassend. Dies gilt sowohl für die Software - 80% ist ungeeignet - als auch für das Datenmaterial, d. h. die Daten- und Informationsbanken.
- Hierdurch sind die Grenzen der Nutzung der elektronischen Informationsdienste in der Schule und im Unterricht festgelegt.
- Nach genauen Überprüfungen stellt der Lehrer fest, daß der Einsatz im Unterricht meist irrelevant ist, bzw. der Arbeitsaufwand, bis Datenbanken und Software einsetzbar sind, unvertretbar hoch ist. Konkret: Lernsoftware ist meist nicht gut durchdacht, die Hersteller geben sich nicht viel Mühe.
- Die Datenbankübungen sind im EDV-(Informatik)-Unterricht kaum von Bedeutung.
- Auch hier fand die aktuelle Qualitätsdiskussion statt, hinter der steckt, daß wir als Kunden nicht zufrieden sind.
- Es wurde auf einige aktuelle Projekte verwiesen, die in nachfolgenden Referaten und in den Workshops vorgestellt wurden.
- Bei der Nutzung von elektronischen Informationen stehen die politischen und nicht die pädagogischen Argumente im Vordergrund.
- Die Euphorie, die sich mit Internet verbindet, sollte sich in Grenzen halten.

Hier wurde eine Vielzahl von kritischen Anmerkungen vorgetragen:

- häufige Staus auf den Datenautobahnen,
- mangelnde Vorkenntnisse zur Nutzung,
- technische Voraussetzungen,
- häufige Irreführungen,
- geringe Qualität vieler Internet-Informationen.

Ein deutsches Schulnetz könnte hier sicherlich wesentliche Verbesserungen bringen.

Veränderungen im Bereich elektronischer Informationsdienste sollten genutzt werden, um Verbesserungen auch für die Aus- und Fortbildung insbesondere in Schulen in der nächsten Zukunft zu erreichen.

Die wichtigsten Veränderungen sind:

- Hochleistungs-Netze durch Glasfaser,
- mehr Konkurrenz durch die Privatisierung der Bundespost,
- kommerzielle Nutzung des Internet,
- Wertschöpfung der Information verändert unser Bewußtsein.

Einige Forderungen wurden in der Diskussion aufgestellt:

- Information ist ein Bürgerrecht. Deshalb sollte niemand durch überteuerte Preise ausgeschlossen werden.
- Die boomende EDV- und Informationsbranche muß sich der Probleme annehmen. Diese sind insbesondere die Erhöhung der Qualität der aus- und fortbildungsrelevanten Daten- und Informationsbanken und die Schaffung solcher Produkte.
- Die Kosten für die Informationsnutzung müssen finanzierbar sein. Die Bibliotheken haben zu wenig Geld für die Informationsnutzung.
- Als spezielle Forderung wurden mehr schulgeeignete elektronische Zeitschriften (Volltext) genannt. Im Bereich der Zeitungsdatenbanken finden wir aber schon schulgeeignete Produkte.
- Die technischen Voraussetzungen schaffen wiederum Barrieren, die abzubauen sind. Es wurde die Software von der Stange gefordert, und diese selbstverständlich im Laienmodus.

Die Daten- und Informationsbankanbieter sollten langfristig denken. Es geht darum, schon in Schulen den Grundstein zu legen für ein Informationsbewußtsein, das sich später, in Studium und Beruf, in konkrete Nachfrage wandelt: "Schüler und Studenten von heute sind die Nutzer von morgen!"

Die Diskussion hat auf verschiedene Punkte aufmerksam gemacht, die es in innovative Lösungsvorschläge umzusetzen gilt. Man sollte nicht klagen, son-

dern konstruktive Lösungsvorschläge erarbeiten. Allein mit Geld läßt sich die Situation der Informationsnutzung in Ausbildungseinrichtungen nicht nur verbessern; Taten sind genauso gefragt.

Anschrift des Verfassers

Prof. Dr. Robert Funk
Humboldt-Universität zu Berlin
Institut für Bibliothekswissenschaft
Clara-Zetkin-Str. 26
10099 Berlin
Tel.: (030) 8236 966; Fax: (030) 8243 052

Friedrich Rost

Erziehungswissenschaftlich relevante elektronische Dienstleistungsangebote[1]

Als Informationsvermittler im Fachbereich Erziehungs- und Unterrichtswissenschaften der Freien Universität Berlin arbeite ich mit Online- sowie mit CD-ROM-Datenbanken. Da ich im Rahmen einer Buchpublikation eine Zusammenstellung von gedruckten und elektronischen Diensten veröffentlicht habe,[2] bot es sich an, Ihnen einen aktuellen Überblick zu geben über erziehungswissenschaftlich relevante elektronische Dienstleistungsangebote. Es handelt sich dabei um Offerten, die in der Regel mit Kosten für den Benutzer verbunden sind. Hier soll jedoch nur am Rande die Kostenseite erörtert werden. Im Vordergrund dieses Referates werden die Inhalte der Datenbanken stehen und die Art der Dienstleistungen, die Datenbankanbieter bzw. Datenbankproduzenten ihren Kunden offerieren. Ich beschränke mich weitgehend auf Literaturdatenbanken, wenngleich es auch Forschungs-, Test- und statistische Datenbasen gibt, die gleichermaßen für Pädagogen und Erziehungswissenschaftler wichtig sein können. Erwarten Sie aber bitte keinen vollständigen Überblick. Der hier zur Rede stehende Markt wandelt sich rasch. Will man sich jedoch elektronischer Datenbanken bedienen, ist eine grundlegende Kenntnis von Möglichkeiten und Grenzen der Recherche in Datenbanken ebenso vonnöten wie eine genauere Kenntnis der Unterschiede zwischen CD-ROM-Datenbanken und Datenbanken auf Großrechnern. Zudem sollte man die für das eigene Fachgebiet relevanten Datenbanken kennen, und sofern man selbst in ihnen recherchieren will, deren jeweilige Feldstruktur und die Struktur der zur Datenbank dazugehörigen Retrievalsprache.

Hinsichtlich der Art der **elektronischen Dienste**, die ich Ihnen hier näher bringen möchte, kann unterschieden werden zwischen:

- Offline-Datenbanken, die nur durch Personal der Institutionen abgefragt werden können, die über die Datenbanken verfügen,

---

[1] etwas erweiterte und nach der Diskussion leicht geänderte Fassung des am 17.11.1994 gehaltenen Vortrags

[2] Friedrich Rost: Auswahlbibliographie. In: Dieter Lenzen (Hrsg.): Erziehungswissenschaft. Ein Grundkurs. Reinbek: Rowohlt, 1994, S. 625-646

- Online-Datenbanken, in denen entweder durch geschultes Personal oder durch den Endnutzer recherchiert werden kann,
- CD-ROM-Datenbanken, die man käuflich erwerben kann bzw. vor Ort in Institutionen wie Bibliotheken oder Online über Computer-Netze benutzen darf, sowie
- Diskettendiensten, also Auszügen aus Datenbanken, die gegen Entgelt verkauft werden.

Möglichkeiten und Grenzen von Datenbankrecherchen

Die Vorteile der *elektronischen* Recherche in Datenbanken gegenüber der herkömmlichen Literaturrecherche in gedruckten Diensten liegt vor allem darin, daß mehrdimensionale Fragestellungen - bei richtiger Anwendung - schnell und präzise bearbeitet werden können, was mittels konventioneller Literatursuche in Katalogen und Bibliographien meist sehr zeitaufwendig ist. Bei der mehrdimensionalen Frage an das System bedient man sich der Boole'schen Operatoren UND, ODER bzw. NICHT. Außerdem kann mittels Maskierungs- bzw. Trunkierungstechnik nach bestimmten Wortstämmen gesucht werden. Anfang und/oder Endungen eines Suchwortes bleiben dabei offen. Allerdings muß erwähnt werden, daß die Links- und Mitten-Trunkierung nicht bei allen Datenbanken (insbesondere nicht bei solchen auf CD-ROM) möglich ist bzw. bei der CD-ROM sehr lange dauert. Doch durch mehrdimensionale Fragestellung und Trunkierungstechnik ist die Suche nach Literaturhinweisen mittels EDV wesentlich effizienter geworden als die in Registern gedruckter Dienste. Zudem müssen die gefundenen bibliographischen Angaben nicht mehr abgeschrieben werden, sondern können bequem ausgedruckt bzw. elektronisch gespeichert werden, z. B. auf Disketten.

Mittlerweile hat die CD-ROM einen solchen Boom erlebt, daß die Online-Datenbanken in der Nutzergunst erheblich in den Hintergrund gedrängt werden. Ich möchte hier jedoch die These vertreten:

> Online-Datenbanken sind umfangreicher, aktueller und zudem schneller durchsuchbar als CD-ROM-Datenbanken oder gedruckte Dienste. Online-Datenbanken sind daher in die Recherche einzubeziehen - zumindest für die sozialwissenschaftliche Arbeit.

Entscheidend für die Zurückhaltung der Nutzer gegenüber Online-Datenbanken dürften zum ersten die Kosten sein, die bei Offline- wie Online-Recherchen nicht im voraus genannt werden können,[3] sofern nicht eine Kostenpau-

[3] Die Kosten für Online-Recherchen z. B. setzen sich in der Regel zusammen aus Entgelten für den Datenbankanbieter (Verbindungszeit, Denkzeit, Dokumentanzeige, Print oder Download), einer Lizenzgebühr für den Datenbankproduzenten, Telekommunikationskosten (alle bisher genannten Kostenarten sind

schale erhoben wird, wie dies in NRW derzeit geschieht. Wo eine kostengünstige Pauschale erhoben wird, ist eine erhebliche Nachfrage nach Online-Datenbankrecherchen zu verzeichnen. In all den anderen Fällen scheuen die Nutzer die vagen Kostenangaben der Hosts bzw. der Informationsvermittler. Oft wird der Preis für die Informationen - auch im nachhinein - als zu hoch empfunden, besonders wenn wenig oder gar nichts gefunden wird, was ja auch eine Information ist (und für eine Forschungsfrage nicht die schlechteste!). Da Online-Recherchen bei Bedienungsfehlern teuer werden können und die Retrievalsprachen komplizierter sind als bei CD-ROM-Datenbanken, die zudem oft kostenlos benutzt werden dürfen, liegen für die meisten Nutzer aus Kosten- und Bedienungsgründen die Vorteile bei der CD-ROM. Darüber werden die Nachteile der CD-ROM-Technik übersehen, wie geringere Aktualität, kleinerer Datenbankumfang und weniger vielseitige Recherchesoftware. In Abbildung 1 (am Ende dieses Beitrags) werden Online- mit CD-ROM-Datenbanken hinsichtlich einiger Kriterien verglichen.

Für Offline, Online wie CD-ROM gilt jedoch, daß es sich in der Regel um Referenz- und keine Volltext-Datenbanken handelt. Referenzdatenbanken geben, ebenso wie Kataloge und Bibliographien, nur Hinweise auf Literatur, die irgendwo eingegeben, inhaltlich erschlossen bzw. angeschafft wurde und die man mit dem Literaturhinweis noch nicht in Händen hält. Hier beginnt die Odyssee der Literaturbeschaffung, ein Problem, das Dienstleister durch Kopierschnelldienste zu lösen trachten.

Grenzen bestehen aber auch hinsichtlich der menschlichen Aufnahmefähigkeit. Angesichts der Literaturmenge steht der Wunsch nach Vollständigkeit der Sammler (hier: der Datenbankproduzenten oder Bibliothekare) dem Wunsch des Nutzers gegenüber, der - häufig unter Zeitdruck stehend - nicht etwa die ganze Literatur durchpflügen, sondern "nur das Relevante" nachgewiesen haben will. So spielen als erste Anhaltspunkte, ob sich die Suche in einer Datenbank lohnen kann, Kriterien eine Rolle wie der Umfang einer Datenbank, eines Katalogs, einer Bibliographie sowie das Sammelgebiet, der Sammlungszeitraum und die Tiefe der Erschließung. Mit Hilfe solcher Kriterien können Rangfolgen oder Leitsysteme aufgestellt werden für Suchanfragen zu den einzelnen Subdisziplinen der Pädagogik, eine Aufgabe, die vielleicht im Rahmen des Modellversuchs FIS Bildung (Implementierung einer Leitstelle) verwirklicht wird.

---

nutzungsabhängig), Portokosten (für Offline-Prints) und die gesetzliche Mehrwertsteuer (vgl. auch Rainer Bucher: Einigermaßen undurchschaubar. Die Preispolitik der Online-Anbieter. In: *Cogito* (1993), H. 4, S. 5-7). Wie man Telekommunikationskosten sparen kann, vgl. Wolfgang Petry: Datenreisen billiger. In: *Cogito* (1994), H. 4, S. 17-22. - Die Kosten für Offline-Recherchen werden meist pauschal - abhängig von der Zahl der Nachweise - erhoben. Auch hier kommen Porto- und/oder Kopierkosten hinzu.

Ein weiteres Problem bei der inhaltlichen Suche besteht durch die manche Nutzer verwirrende Vielfalt an unterschiedlichsten Ordnungssystemen wie Systematiken, Klassifikationen, Thesauri, Schlagwortlisten.[4] Doch überall, wo Denkinhalte über Sprache symbolisch vermittelt werden (müssen), gibt es Quellen des inhaltlichen Mißverständnisses und des wissenschaftlichen Dissenses in terminologischen Fragen.[5] Doch sind dies generelle Schwierigkeiten der inhaltlichen Erschließung und keine durch elektronische Dienste hervorgerufenen. Grundsätzlich sollte man sich jedoch mit den jeweiligen inhaltlichen Ordnungsmitteln vertraut machen, um deren Struktur und Gliederungsfähigkeit kennenzulernen.[6]

Hinsichtlich der inhaltlichen Recherche bieten Datenbanken auch dabei erhebliche Vorteile gegenüber Katalogen und gedruckten Diensten, weil man mit einer umfassenden Suche beginnen und bei zu hohen Trefferzahlen die Fragestellung vorsichtig thematisch einschränken kann. Dabei muß allerdings berücksichtigt werden, daß sozialwissenschaftliche internationale Datenbanken meist erst nach 1960 begonnen wurden, in Deutschland oft erst nach 1975, und daß bisher erst für wenige Fachgebiete auch ältere Literatur nachträglich eingegeben wurde. Doch ich bleibe dabei:

> Für jede sorgfältige Literatursuche, zumal für ein wissenschaftliches Projekt, sollten alle Möglichkeiten elektronischer wie konventioneller Recherche eingesetzt werden, um die bisherige Literatur und Forschung zur Kenntnis zu nehmen und Doppelarbeit zu vermeiden.

Offline-Datenbanken

Nach diesen Vorbemerkungen will ich überleiten zu den Offline-Datenbanken, die nur durch Personal der jeweiligen Institution abgefragt werden können, weil derzeit kein Online-Zugang möglich ist. Dankenswerterweise hat der Modellversuch Fachinformationssystem Bildung 1993 "ein kommentiertes Verzeichnis von Fachinformationseinrichtungen und ihren Dienstleistungen" herausgebracht, das durch eine Vielzahl von Registern mustergültig inhaltlich erschlossen ist, so daß ich wegen der Fülle der Einrichtungen (über 40) und auch der Vielzahl der Inhouse-Datenbanken einfach auf dieses preis-

---

[4] Vgl. dazu Klaus-Peter Horn/Lothar Wigger (Hrsg.): Systematiken und Klassifikationen in der Erziehungswissenschaft. - Weinheim: Deutscher Studien Verlag, 1994.

[5] Vgl. dazu meine Position: Friedrich Rost: Pädagogische Bibliographien, Thesauri und Register. In: Klaus-Peter Horn/Lothar Wigger (Hrsg.): a.a.O., S. 191-214.

[6] Hierzu merkte Prof. Dr. Thomas Feuerstein in der sich an den Vortrag anschließenden Diskussion an, daß dies ratsamerweise bei ersten Recherchen an Literatur geschehen sollte, die man sehr gut kennt, um die Umsetzung des Inhalts in die Ordnungssysteme zu überprüfen.

werte Buch verweisen möchte.[7] Darüber hinaus muß die von Heinz Ziegler und Peter Diepold 1993 erarbeitete Broschüre für die "Beschaffung von Literaturinformationen zur beruflichen Bildung" genannt werden, die kostenlos vom Bundesministerium für Bildung und Wissenschaft abgegeben wird.[8] Beide Wegweiser sollten gepflegt und fortgeschrieben werden, so daß einerseits aktuelle Datenbanken vorhanden sind, zum anderen die Drucklegung neuer Auflagen rasch erfolgen kann.

Abgesehen von eventuell notwendigen Rückfragen bezüglich der Kostenübernahme und vor allem zur genauen Eingrenzung der Themenstellung für die Recherche sind nach meinen Erfahrungen Offline-Datenbankanfragen sehr hilfreich, weil

1. das Personal einer Institution die eigene Datenbank am besten kennt,
2. die nachgewiesene Literatur oft auch im Hause ist, weil die Informations- und Dokumentationsstellen meist die ausgewerteten Materialien in einer institutseigenen Bibliothek bereithalten,
3. die Kosten für die erbrachten Dienstleistungen meist geringer sind als bei den Online-Diensten, an denen Produzent und Host (zu ihrer Kostendeckung) "verdienen" müssen.

Online-Datenbanken

Online-Dienste werden entweder von den Datenbankproduzenten selbst zur Verfügung gestellt und/oder von den sogenannten "Hosts", das sind Großrechenzentren, die die Datenbanken und ihren Service gegen Entgelt anbieten. Die Dienstleistungspalette reicht

- von der Bereitstellung der Datenbanken für Recherchen, die man selbst von einem PC aus durchführt, so man einen Datenfernübertragungsanschluß hat,
- über Recherchen, die Mitarbeiter des Datenbankproduzenten bzw. -anbieters in Ihrem Auftrag durchführen und deren Ergebnisse man dann zugeschickt erhält,
- bis hin zu sogenannten Profildiensten (Selective Dissemination of Information = SDI), das sind Suchläufe zu einem bestimmten Themenbereich nach

---

[7] Fachinformationssystem Bildung: Dokumentationseinrichtungen im Bereich Bildung - Bundesrepublik Deutschland 1992. Ein kommentiertes Verzeichnis von Fachinformationseinrichtungen und ihren Dienstleistungen. Berlin: Verlag für Wissenschaft und Bildung, 1993

[8] Heinz Ziegler/Peter Diepold: Beschaffung von Literaturinformationen zur beruflichen Bildung. Ein erster Wegweiser. Hrsg. vom Bundesministerium für Bildung und Wissenschaft. Bonn, 1993. - Erhältlich über die Pressestelle des Ministeriums, Postfach 200108, 53175 Bonn

jedem Update der Datenbank, bei denen fortlaufend Literatur zu Ihrem Gegenstand gesucht wird, bis der Dauerauftrag von Ihnen gelöscht oder geändert wird.

Der Vorteil von Datenbankrecherchen auf Großrechnern besteht darin, daß z. T. sehr große Datenbanken ohne Segmentierung, d. h. unzerstückelt, durchsucht werden können (vgl. auch Abbildung 1). Mittlerweile kann man bei einigen Hosts sogar mit einer Fragestellung mehrere Datenbanken zugleich im Superbase-Modus durchsuchen. Hier ist eine gegenläufige Entwicklung zur Suche auf CD-ROM eingetreten. Während bei großen Datenbanken die Daten nicht mehr auf eine CD-ROM passen, so daß zwei oder mehr Scheiben nacheinander durchsucht werden müssen, werden auf Großrechnern z. T. fünf und mehr Datenbanken im Gigabytebereich zugleich nach einer Fragestellung durchsucht.[9] CD-ROM-Versionen von Online-Datenbankinhalten unterscheiden sich hinsichtlich Datenbankumfang, Struktur und Zahl der Datenbankfelder und der Retrievalsoftware. Abbildung 2 stellt die Daten der CD-ROM "DNB aktuell" denen von BIBLIODATA gegenüber, der Online-Datenbank der Deutschen Bibliothek.

Erziehungswissenschaftlich relevante Online-Datenbanken mit einem hohem Anteil an deutschsprachiger Literatur

Leider fehlt derzeit eine originär erziehungswissenschaftliche Online-Datenbank mit Referenzen auf deutschsprachiges Schrifttum, wie sie im 1. IuD-Programm der Bundesregierung (1974 - 1977) geplant war.[10] Das darin vorgesehene Fachinformationszentrum Bildung wurde nie gegründet. Auch der Antrag und die Konzeption des Modellversuchs Fachinformationssystem Bildung sehen keine zentrale Online-Datenbank mehr vor.[11] Deshalb ist die Suche in allgemeinbibliographischen Datenbanken sowie solchen der Nachbardisziplinen erforderlich, die z. T. über erhebliche Referenzzahlen zu erziehungswissenschaftlich/pädagogischen Beständen verfügen.

BIBLIODATA. Gespeichert sind die bibliographischen Angaben von Büchern, Zeitschriften und Karten, die die Deutsche Bibliothek als Archivbi-

[9] Ob das angesichts der unterschiedlichen inhaltlichen Erschließung immer sinnvoll ist, ist eine diskussionswürdige Frage.

[10] Vgl. dazu: Bundesministerium für Forschung und Technologie: Programm der Bundesregierung zur Förderung der Information und Dokumentation (IuD-Programm) 1974-1977. Bonn 1975.

[11] Vgl. dazu: Fachinformationssystem Bildung: Zur Situation der Fachinformation im Bereich Bildung. Eine Bestandsaufnahme bei IuD-Einrichtungen in der Bundesrepublik Deutschland mit Schlußfolgerungen für die Konzeption des Fachinformationssystems (FIS) Bildung. Eschborn, 1992.

bliothek sammelt. Grundlage für Bibliodata ist die Deutsche Bibliographie mit ihren Reihen A (= Veröffentlichungen des Verlagsbuchhandels), B (= Veröffentlichungen außerhalb des Verlagsbuchhandels = Graue Literatur), C (= Karten), H (= Hochschulschriften), N (= Kommende Neuerscheinungen). Der Datenbestand liegt derzeit bei ca. 2,2 Millionen Dokumenten, die zwischen 1972 und heute publiziert wurden. Eine Suche ist mit den Boole'schen Operatoren in 26 Feldern möglich, eine inhaltliche Suche z. B. nach Sachgruppen (von 36 auf 65 Sachgruppen erweitert [ab 1982] [Erziehung, Bildung, Unterricht = 10a bzw. 22]), nach Schlagwörtern, über die ca. 2/3 der Dokumente verfügen, oder durch trunkierte oder nichttrunktierte Stichwortsuche im Basic Index. Eine Besonderheit stellt in einigen Feldern des Index die Suche mit Proximity-Operatoren dar, mit deren Hilfe eine Verknüpfung von Wörtern möglich ist, die in bestimmter Beziehung zueinander stehen sollen. BIBLIODATA ist über STN und das DBI erreichbar, einen Auszug der Datenbank gibt es auf CD-ROM (vgl. Abbildung 2).

ELIS. Das Erlangen Library Information System, abgekürzt ELIS, enthält den Gesamtbestand der Literatur der Universitätsbibliothek Erlangen-Nürnberg ab 1982, teilweise auch früher (derzeit über 1,1 Millionen Bände). Die UB Erlangen-Nürnberg betreut im Auftrag der Deutschen Forschungsgemeinschaft das Sondersammelgebiet Bildungsforschung und verfügt dementsprechend über einen hohen Anteil nationaler wie internationaler pädagogischer Literatur, und zwar ca. 400.000 Bände. In ELIS ist aber nur der Bestand, der ab 1982 angeschafft wurde, nach Schätzung von Herrn Theuerkauf, UB Erlangen-Nürnberg, ca. 60.000 Bände. Neben den vollständigen Katalogdaten der Monographien und Sammelwerke sind Schlagwörter zu den Büchern vergeben, nach denen ebenso gesucht werden kann wie nach Titelstichwörtern. Es erfolgt keine Aufnahme und inhaltliche Erschließung von Aufsätzen! Die Datenbank ist über Internet, WIN, Datex-P und Telefon erreichbar und derzeit noch kostenfrei. Ein Teil des Bestandes ist auf der CD-ROM des FIS Bildung (vgl. auch Abbildung 3).

FORIS. Beim Forschungsinformationssystem Sozialwissenschaften (FORIS) handelt es sich um eine Forschungsdokumentation, die seit 1978 über 40.000 Projekte aus Deutschland, Österreich und der Schweiz enthält, und die dreimal im Jahr ergänzt wird durch Neuaufnahmen aufgrund der schriftlichen Erhebungen durch das Informationszentrum (IZ) Sozialwissenschaften. Da die Angaben freiwillig sind, melden allerdings viele Wissenschaftler ihre Projekte nicht oder nicht sofort, so daß mit erheblichen Lücken zu rechnen ist. Die Nachweise enthalten ausführliche Angaben zum Thema, zu Inhalt und methodischen Vorgehen, zu Verfahren der Datengewinnung sowie formalen Angaben der Institutionenanbindung, zu Namen der Mitarbeiterinnen und Mitarbeiter, der Laufzeit und Finanzierung. Verschlagwortet werden die Projekte nach der Schlagwortliste des Informationszentrums (IZ) Sozialwissenschaften. Jährlich kommen ca. 5.000 Projekte hinzu. Der Anteil erziehungswissenschaftlicher Projekte an der Datenbank wird vom IZ mit ca. 20 %

angegeben. FORIS kann über die Hosts STN, DIMDI, GBI online recherchiert werden oder auf der CD-ROM WISO III (vgl. auch Abbildung 4).

MATHDI. Die Datenbank Mathematikdidaktik, abgekürzt MATHDI, enthält Literaturhinweise auf das deutschsprachige sowie relevantes anderssprachiges Schrifttum zum Mathematikunterrricht und seinen Grenzbereichen, so auch zur Informatik und informationstechnischen Bildung. Ausgewertet werden Bücher und Zeitschriftenaufsätze, Graue Literatur, audiovisuelle Medien, Spiele und anderes. Eine Dokumentationseinheit setzt sich zusammen aus der bibliographischen Angabe, einer Notation der Klassifikation, freien Schlagwörtern und meist einem Abstract in Deutsch und Englisch. MATHDI umfaßt über 57.000 Dokumentationseinheiten (DE) seit 1976 und wird zweimonatlich aktualisiert. MATHDI ist online über STN verfügbar. Einen Auszug von MATHDI gibt es auch auf CD-ROM (vgl. auch Abbildung 5).

PSYNDEX. Der psychologische Index, kurz: PSYNDEX, umfaßt vorwiegend deutschsprachige psychologische Literatur, die die Zentralstelle für psychologische Information und Dokumentation (ZPID) in Trier formal erfaßt und inhaltlich auswertet. PSYNDEX lehnt sich stark an sein amerikanisches Vorbild PsycINFO an. So ist PSYNDEX auch mit englischsprachigen Deskriptoren durchsuchbar und zum großen Teil mit englischsprachigen Abstracts versehen. Neben Monographien, Sammelwerken und Zeitschriften werden Hochschulschriften, Aufsätze und audiovisuelle Medien nachgewiesen. PSYNDEX umfaßt über 90.000 Dokumente aus dem Zeitraum 1977 bis heute und wird monatlich aktualisiert. Der Anteil an erziehungsrelevanter Literatur, nicht nur zur pädagogischen Psychologie, liegt nach meinen Datenbankrecherchen mit bestimmten Wortstämmen im Deskriptorfeld[12] bei ca. 33 %. PSYNDEX ist über DIMDI und GBI online sowie auf CD-ROM verfügbar (vgl. auch Abbildung 6).[13]

PSYTKOM ist eine Datenbank deutschsprachiger psychologischer und pädagogischer Tests der ZPID in Trier. Die deutschsprachige Testbeschreibung besteht aus bibliographischen Angaben, Ausführungen über das Testkonzept und die Testkonstruktion, über Gütekriterien, Durchführungs- und Auswertungsmodalitäten. PSYTKOM weist außerdem Forschungs- und Sekundärliteratur zu diesen Tests nach sowie einen zusammenfassenden kritischen Kommentar. Die Tests sind mit Hilfe des amerikanisch-deutschen Thesaurus verschlagwortet, so daß auch mit englischen Deskriptoren gesucht werden

---

12 z. B. "?ausbild?", "?bildung", "?didakti?", "?erzieh?", "?hilf?", "?jugend?", "?kind?", "?lehr?", "?lern?", "?paedagog?", "?schul?", "?sozialisat?", "?unterricht?" und anschließenden rechnerischen Bereinigungen um inhaltlich nicht dazugehörige Mengen und Überschneidungen (zum Beispiel "Schulkind")

13 Welche Unterschiede bei PSYNDEX die Online- und die CD-ROM-Version hinsichtlich der Recherchesoftware und Suchmöglichkeiten aufweisen, darüber berichtete Frau Labouvie im Workshop 2 dieser Tagung; vergleiche ihren Beitrag in diesem Band.

kann. Die Datenbank umfaßt ca. 2.500 Tests seit 1945. Eine Aktualisierung erfolgt halbjährlich, jährlich kommen ca. 120 Tests hinzu. PSYTKOM ist über DIMDI online verfügbar (vgl. auch Abbildung 7).

SOLIS. Das sozialwissenschaftliche Literaturinformationssystem SOLIS des IZ Sozialwissenschaften umfaßt über 160.000 Dokumentationseinheiten seit 1945, jedoch im Schwerpunkt den Zeitraum von 1977 bis heute. Der Bestand besteht zu ca. 42 % aus Zeitschriftenaufsätzen, zu 45 % aus Büchern und Sammelwerksbeiträgen und zu 10 % aus Grauer Literatur. 88 % der Nachweise haben ein Abstract, oft auch in Englisch. Neben den üblichen bibliographischen Angaben sind die Dokumentationseinheiten deutsch und englisch verschlagwortet nach der Schlagwortliste des IZ. Die Datenbank hat einen Zuwachs von ca. 15.000 Dokumenten jährlich und wird monatlich aktualisiert. Der Anteil an pädagogisch interessierender Literatur liegt nach strenger klassifikatorischer Zuordnung bei 4 %, nach meinen Berechnungen zu den in Fußnote 6 genannten Wortstämmen im Feld CT (= Controlled Terms) bei ca. 33 %. SOLIS ist online über DIMDI, STN oder GBI verfügbar oder auszugsweise über die CD-ROM WISO III (vgl. auch Abbildung 8).

SOMED, die Datenbank zur Sozialmedizin des Instituts für Dokumentation und Information Sozialmedizin, öffentliches Gesundheitswesen (idis) in Bielefeld, enthält eine große Menge von Dokumentennachweisen zur Gesundheitserziehung und zur Sucht- sowie Umweltproblematik. Die Datenbank umfaßt seit 1978 über 277.000 Dokumentationseinheiten aus ca. 700 Fachzeitschriften, aus Büchern, Sammelwerksbeiträgen, Dissertationen, Grauer Literatur. Nach meiner Berechnung sind ca. 25 % des Bestandes, also etwa 75.000 Dokumente, pädagogisch relevant. 62 % der Dokumentationseinheiten sind mit einem Abstract versehen. Der Bestand beinhaltet zu je 45 % Referenzen auf deutsch- und englischsprachige Literatur. Der Rest verteilt sich auf andere Sprachen. Ca. 85 % der Nachweise kann über den idis-Literaturschnelldienst in Kopien bezogen werden. Der jährliche Zuwachs beträgt zwischen 15.000 und 20.000 Dokumentationseinheiten, die Datenbank wird monatlich aktualisiert. SOMED ist nur über DIMDI online verfügbar (vgl. auch Abbildung 9).

Relevante Online-Datenbanken mit überwiegend fremdsprachiger Literatur

BRITISH EDUCATION INDEX heißt neben dem gedruckten Dienst auch die Datenbank, die Dokumentennachweise aus 200 - 250 ausgewählten pädagogischen Zeitschriften ab 1972 verzeichnet, und zwar über 55.000. Darüber hinaus enthält sie über 10.300 Hochschulschriften des BRITISH EDUCATION THESES INDEX (BETI). Verschlagwortet ist der Bestand nach dem British Education Thesaurus. Die Datenbank ist über DIALOG online verfügbar. Neu angekündigt bei DIALOG ondisc wird "International ERIC", eine CD-ROM-Datenbank, die den Australian, den British sowie den Canadian Education Index umfassen soll mit Beständen ab 1976 (vgl. auch Abbildung 10).

ERIC, die Datenbank des amerikanischen Educational Resources Information Center ERIC, ist auch Grundlage für die gedruckten Dienste "Current Index to Journals in Education (CIJE)" und "Resources in Education (RIE)". Die Datenbank enthält Literaturhinweise aus mehr als 700 vorwiegend amerikanischen pädagogischen Fachzeitschriften und Nachweise von Forschungsreports, also Grauer Literatur, zum Zeitraum 1966 bis heute. Derzeit sind es ca. 700.000 Dokumentationseinheiten. Es besteht die Möglichkeit, viele Dokumente per online ordering in Kopie zu bestellen. Die Datenbank ist online über DIALOG sowie im Internet über ASK ERIC[14] verfügbar, aber auch auszugsweise als CD-ROM (vgl. auch Abbildung 11).

FRANCIS Sciences de L'éducation. Die Datenbank des Institut de L'information Scientifique et Technique ist seit 1972 computerisiert, enthält über 110.000 Dokumentationseinheiten zur Pädagogik und wächst jährlich um 3.500 Dokumente. 91 % davon sind Zeitschriftenartikel. Die Verschlagwortung erfolgt in Französisch und Englisch. 83 % der Dokumentationseinheiten umfassen ein Abstract. Die Datenbank ist online über QUESTEL recherchierbar (vgl. auch Abbildung 12).

PsycINFO enthält alle Zitate aus dem gedruckten Dienst "Psychological Abstracts" sowie psychologische Literatur aus "Dissertation Abstracts". Es werden Beiträge aus ca. 1.500 Zeitschriften aufgenommen, ebenso Bücher und Dissertationen. Seit 1969 sind über 909.000 Dokumentationseinheiten gespeichert, jährlich kommen ca. 30.000 hinzu. 77 % der Dokumentationseinheiten sind mit einem Abstract versehen. Die Hauptsprache der Dokumente ist Englisch. Deutsche Literatur ist selten, da PSYNDEX seit 1977 - in Anlehnung an PsycINFO - das deutschsprachige psychologische Schrifttum auswertet. Die Datenbanken sind zueinander kompatibel, so daß mit den gleichen Deskriptoren des amerikanisch-deutschen Thesaurus in beiden Daten-

[14] Adresse: http://ericiv.sunsite.syr.edu

banken gesucht werden kann. Für PsycINFO steht bei DIMDI wahlweise eine Benutzerführung in Deutsch und Englisch zur Verfügung. Einen Datenbankauszug gibt es auch unter der Bezeichnung PSYCLIT auf CD-ROM (vgl. auch Abbildung 13).

Social SCISearch entspricht inhaltlich dem "Social Science Citation Index", der alle Gebiete der Sozial- und Geisteswissenschaften umfaßt. Über 2,5 Millionen Referenzen seit 1973 umfaßt diese Datenbank. Ausgewertet werden jährlich etwa 120.000 Aufsätze aus ca. 1.500 meist englischsprachigen Zeitschriften sowie auszugsweise aus weiteren 2.400 Zeitschriften des "Science Citation Index". Der Anteil der Nachweise auf Dokumente in deutscher Sprache beträgt 5 % (über 118.000 Dokumentationseinheiten). Nachgewiesen werden Autoren, Titel, bibliographische Angaben, Institutsadressen sowie in einem Drittel des Bestandes seit 1992 Abstracts. Als Besonderheit des SocialSciSearch gilt, daß die Literaturreferenzen der nachgewiesenen Dokumente in Kurzform aufgenommen sind. Mit der Annahme, daß Literatur nur angeführt wird, wenn sie inhaltlich eine Bedeutung für das Thema hat, kann man, also weiter suchen, wer dieses Schrifttum auch zitiert hat. Die Datenbank ist über DIALOG online verfügbar oder auszugsweise über CD-ROM. Wen nur die neueste Literatur der letzten 12 Monate interessiert, sollte online die Teildatendank CURRENT CONTENTS/SSCI anwählen (vgl. auch Abbildung 14).

Sociological Abstracts (SA) entspricht inhaltlich der gedruckten Ausgabe, umfaßt aber auch Dokumentationseinheiten aus "SOCIAL PLANNING/ POLICY & DEVELOPMENT ABSTRACTS (SOPODA)" sowie "International Reviews of Publications in Sociology (IRPS)". Die Datenbank beinhaltet über 385.000 Zitate ab 1963, der jährliche Zuwachs liegt bei ca. 19.000 Dokumenthinweisen, davon 54 % mit Abstracts. Ausgewertet werden über 1.600 Zeitschriften sowie Bücher, Konferenzberichte, Dissertationen. Die Datenbank wird fünfmal im Jahr aktualisiert und ist über DIMDI und DIALOG online recherchierbar oder auszugsweise über die CD-ROM SOCIOFILE (vgl. auch Abbildung 15).

Bei all den genannten Datenbanken ist gegen Entgelt auch das Downloading der Referenzen für die eigene Arbeit erlaubt, so daß die Daten in ein eigenes Literaturverwaltungsprogramm geladen werden können, wie es Peter Diepold in seinem Vortrag kurz erläutert[15] und Ortwin Peske auf dem GIB-Frühjahrs-Workshop 1994 anhand von "LIDOS-Download 2.0" demonstriert haben. Die Abbildung 16 faßt noch einmal die Datenbanken nach den jeweiligen Hosts zusammen; weitere über die Hosts erreichbare Datenbanken der pädagogischen Grenzgebiete sind in Klammern lediglich genannt, hier jedoch nicht beschrieben worden.

[15] vgl. Peter Diepold: Elektronische Informationsdienste für Wissenschaftler, in diesem Band S. 15ff.

Diskettendienste

Einige Datenbankproduzenten, wie beispielsweise das IZ Sozialwissenschaften, das Deutsche Institut für Internationale Pädagogische Forschung (DIPF), das Deutsche Zentralinstitut für soziale Fragen (DZI) oder die religionspädagogischen Institute (Comenius-Institut, Münster; Katholische Religionspädagogische Dokumentationsstelle, München) bieten Diskettendienste an. Hierbei handelt es sich um Datenbankauszüge entweder zu einem Themenbereich oder zu dem Bestand eines Jahres. Die Dokumentation kann auf den eigenen PC geladen und mit Hilfe eines Recherchemoduls durchsucht werden.

Literaturbeschaffung

Engpaß nach all den gefundenen Literaturreferenzen ist - wie eingangs schon erwähnt - die Literaturbeschaffung. Hieran arbeiten die Hosts mit erheblichem Aufwand zumindest für den Bereich der Aufsatzliteratur. Ich verweise zudem auf das Projekt JASON-NRW, das in Nordrhein-Westfalen unter Beteiligung großer Bibliotheken des Landes NRW läuft. Hinsichtlich der Monographien und Sammelwerke ist es ratsam, sich bei Zentralbibliotheken oder Verbundkatalogen (z. B. online beim GBI "MONO") zu erkundigen, wo das entsprechende Buch vorrätig ist, so lange wir noch auf öffentlich zugängliche OPAC-Kataloge warten müssen oder auf elektronische Bibliotheken mit Volltextdatenbanken.

Schlußbemerkungen

Wie zu sehen war, gibt es bereits eine - vielleicht verwirrende - Fülle von Datenbanken, die auch für erziehungswissenschaftliche Fragestellungen genutzt werden können. Abgesehen von der Kostenklippe, die m. E. durch eine Fehlentscheidung der Bundesregierung in der Fachinformationspolitik verursacht ist,[16] gibt es allerdings noch erhebliche Informationslücken bei den Pädagogen und Erziehungswissenschaftlern hinsichtlich der elektronischen Dienstleistungsangebote. Bei der neuen Studentengeneration und dem wissenschaftlichen Mittelbau erlebe ich durchweg eine erfreuliche Aufgeschlossenheit und Unbefangenheit im Hinblick auf die elektronische Recherche. So glaube ich, daß sich die Arbeit für den Modellversuch FIS Bildung und die Gesellschaft Information Bildung letztlich auch lohnen wird,

[16] Vgl. hierzu ausführlicher: K. A. Stroetmann: Fachinformationsdienste und gesellschaftlich produziertes Wissen: Ökonomische, gesellschafts- und informationspo-

indem die Infrastrukturverbesserungen sich auch in einer Qualitätssteigerung der erziehungswissenschaftlichen Lehre, Forschung und Literatur niederschlagen werden und damit vielleicht auch in der pädagogischen Ausbildung und Praxis.

Anschrift des Verfassers:

Dr. Friedrich Rost
Freie Universität Berlin
FB Erziehungs- und Unterrichtswissenschaften
SE Information & Datenverarbeitung
Arnimallee 10
14195 Berlin
Tel: (030) 838-5888; Fax: -5889

litische Fragestellungen und Perspektiven. In: *Nachrichten für Dokumentation* 41(1990), S. 343-348.

Abbildung 1a: Online und CD-ROM im Vergleich - I

Kriterium	Online	CD-ROM
Datenbankgröße:	> 1 GByte	vorzugsweise < 600 MB, darüber hinaus Segmentierung nötig
Datenbankstrukturen von DB:	sehr unterschiedlich	sehr unterschiedlich
Aktualität:	wö. - 1/2-jährl. Update	kein - 12 x jährl. Update
Kosten:	nutzungsabhängig, schwer im vorhinein kalkulierbar	überschaubare, fixe Kosten
Geräte-Aufwand:	einfacher PC, Modem; Telefonleitung, darüber mehrere Hosts erreichbar	leistungsfähiger PC, möglichst Farbmonitor, CD-ROM-Laufwerk(e)
Erstinstallation erfordert:	Datex-P- und DOS-Kenntnisse, Anmeldung bei Hosts	Hard- und Software-Kenntnisse, Verfügbarkeit der CD-ROMs
Materialarten:	überwiegend Textinformation (Grafiken u. Bilder nur mit hostabhängiger Zusatz-Software)	Textinformation, doch auch Speicherung u. Wiedergabe von Ton, Grafiken u. Bildern auf CD-ROM möglich
Verfügbarkeit:	mehrere Benutzer können von verschiedenen Geräten aus problemlos gleichzeitig arbeiten	mehrere Benutzer können relativ problemlos von verschiedenen Geräten aus gleichzeitig arbeiten (wenn PC-Netzwerk vorh.)

Abbildung 1b: Online und CD-ROM im Vergleich - II

Kriterium	Online	CD-ROM
Such- und Zugriffszeiten:	auch bei großen Datenmengen schnell (abhängig von Größe u. Belastung des Systems)	durch physikalische Eigenschaften der CD-ROM begrenzt, deutlich langsamer als Online
Retrieval-sprachen:	3-5 versch. Retrieval-sprachen decken Bereich der Online-Datenbanken ab	einige führende Retrieval-Systeme, doch auch zahlreiche "Exoten"
Qualität und Möglichkeiten der Retrieval-Software:	sehr ausgereift; sehr viele Möglichkeiten, z. B. Kontextoperationen	meist weniger vielseitig, einfacher
Lernaufwand für Retrieval-Software:	2-3 Tage	2 - 3 Stunden
Hilfe-Funktionen:	sehr beschränkte Online-Hilfefunktion; Help desk	i. d. R. ausführliche, kontextorientierte Hilfe-Funktionen
Technik, Anwenderkreis: (Fazit 1):	Erprobte Technik im Stadium der Reife, für geübte Anwender	Junge Technik mit boomartigem Erfolg; auch als Einstieg für ungeübte Benutzer
bei guten Kenntnissen der Datenbankstruktur und Retrievalmöglichkeiten (Fazit 2):	bessere Ergebnisse, weil aktuellere, unsegmentierte Datenbestände mit ausgefeilterer Retrievalsoftware schneller durchsucht werden können	ordentliche Ergebnisse, vielleicht nicht ganz so aktuell; ältere Bestände vielleicht gar nicht im Bestand der CD-ROM

(In Anlehnung an die Tabelle von Wolfgang Zick. In: Cogito [1990], H.1, S. 12)

Abbildung 2: Vergleich BIBLIODATA - DNB aktuell

BIBLIODATA

Datenbestand (1.11.94):	**ca. 2,2 Mill. Dokumente (nur Bücher und Zeitschriften, keine Aufsätze!)**
abgedeckter Zeitraum:	**ab 1972 bis heute**
Update:	**wöchentlich ca. 10.000 Dokumente**
Datenbankstruktur:	**26 Felder**
Besonderheiten:	**Im Basic Index ist die Suche mit Proximity-Operatoren möglich**

DNB aktuell (2 CD-ROM)

Datenbestand (1.9.94):	**ca. 700.000 Dokumente (nur Bücher und Zeitschriften, keine Aufsätze!)**
abgedeckter Zeitraum:	**ab 1986 bis heute (1. CD: 1986 - 1992, 2. CD: 1991ff.); dadurch teilweise Überschneidung des Bestands**
Update:	**4 x im Jahr**
Datenbankstruktur:	**17 Felder**

(Quelle: Blue Sheet BIBLIODATA von STN; CD-ROM-Anleitung DNB)

Abbildung 3

Name der DB	Erlangen Library Information System ELIS (IMMD 4)
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☐
Fachgebiet(e)	alle Fachgebiete der Universität Erlangen-Nürnberg
Zahl der DE insgesamt	über 1,1 Millionen
abgedeckter Zeitraum	ab 1982, z. T. auch früher, wenn ab 1982 angeschafft
dokumentierte Materialien	Monographien, Serien, Sammelwerke, Graue Literatur, Zeitschriften, **keine Aufsätze!**
Abstracts	keine
Sprachanteile	k. A.
für die EWI relevant	ca. 60.000 Bände Sondersammelgebiet Bildungsforschung
Hersteller/Anbieter	UB Erlangen-Nürnberg
Kosten	derzeit kostenfrei
Host(s)	direkt über DATEX-P, WIN, INTERNET CD-ROM: Lit.dok. Bildung (FIS Bildung)

Abbildung 4

Name der DB	Forschungsdokumentation FORIS
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☐
Fachgebiet(e)	Forschungsprojekte in den Sozialwissenschaften aus Deutschland, Österreich und der Schweiz
Zahl der DE insgesamt	ca. 40.000 / Zuwachs jährl. ca. 5.000
abgedeckter Zeitraum	ab 1978 / Update: 3 x im Jahr
dokumentierte Materialien	Projekte aus ausgewerteten Fragebögen (jährliche Erhebung)
Abstracts	ca. 93 % mit inhaltl. Kurzdarstellung des Projekts
Sprachanteile	vorwiegend deutsch
für die EWI relevant	knapp 20 % (Auskunft IZ)
Hersteller/Anbieter	IZ Sozialwissenschaften
Kosten	abhängig vom Host
Host(s)	STN, DIMDI, GBI / CD-ROM: WISO III

Abbildung 5

Name der DB	Mathematikdidaktik MATHDI
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☐
Fachgebiet(e)	Mathematikdidaktik & -unterricht sowie dessen Grenzbereiche, z. B. Informatik, informationstechnische Bildung
Zahl der DE insgesamt	über 57.000 / Jährl. Zuwachs: 4.000
abgedeckter Zeitraum	seit 1976 / Update: alle 2 Monate
dokumentierte Materialien	Zeitschriftenaufsätze und Serien (50%), Bücher, insbesondere Lehr- und Schulbücher (30%), Graue Literatur, audiovisuelle Medien, Spiele u. a.
Abstracts	ja, zum Anteil k. A.
Sprachanteile	k. A.
für die EWI relevant	z. B. Unterrichtsforschung, Schulbuchforschung
Hersteller/Anbieter	FIZ Karlsruhe sowie Zentrum f. Didaktik d. Mathematik
Kosten	195 DM/h & 1,50 DM pro Dokument
Host(s)	STN / Auszug auf CD-ROM MATHDI

Abbildung 6

Name der DB	Psychologischer Index PSYNDEX
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☐
Fachgebiet(e)	Gesamte Psychologie einschl. psycholog. relevanter Aspekte aus Nachbardisziplinen wie Psychiatrie, Soziologie, Erziehungswissenschaft u. a.
Zahl der DE insgesamt	über 90.000 / Jährl. Zuwachs: 8.000
abgedeckter Zeitraum	seit 1977 Update: 1/4-jährlich
dokumentierte Materialien	Zeitschriften- und Sammelwerkaufsätze, Monographien, Hochschulschriften, Graue Literatur, audiovisuelle Medien
Abstracts	in deutsch und englisch
Sprachanteile	fast ausschließlich deutsche Quellen
für die EWI relevant	ca. 33 %
Hersteller/Anbieter	ZPID, Universität Trier
Kosten	hostabhängig
Host(s)	DIMDI, GBI / CD-ROM: PSYNDEX

Abbildung 7

Name der DB	Testdatenbank PSYTKOM
angebot. Form(en)	online ☒ CD-ROM ☐ Diskettendienst ☐
Fachgebiet(e)	Psychologie, Pädagogik
Zahl der DE insgesamt	ca. 2.500 / Jährl. Zuwachs: ca. 120
abgedeckter Zeitraum	ab 1945 Update: halbjährlich
dokumentierte Materialien	psychologische und pädagogische Tests
Abstracts	ja
Sprachanteile	ausschließlich deutschsprachige Materialien
für die EWI relevant	ja
Hersteller/Anbieter	ZPID Trier
Kosten	10 DM/h & 0,75 DM pro Dokument
Host(s)	DIMDI

Abbildung 8

Name der DB	Sozialwissenschaftliches Literaturinformationssystem SOLIS
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☒
Fachgebiet(e)	Soziologie, Sozialpolitik, Sozialgeschichte, Bevölkerungsforschung, Arbeitsmarkt- und Berufsforschung, Kommunikations-wissenschaft, Sozialpsychologie
Zahl der DE insgesamt	über 160.000 / Zuwachs jährl. ca. 15.000
abgedeckter Zeitraum	seit 1977, Monographien auch davor (ab 1945) / Update: monatlich
dokumentierte Materialien	42 % Zeitschriftenaufsätze, 13 % Monographien, 32,5 % Sammelwerke und Beiträge aus Sw. 10 % Graue Literatur
Abstracts	88 % der DE, ein Teil davon auch in Englisch
Sprachanteile	fast ausschl. deutschsprachige Materialien
für die EWI relevant	nach strenger Klassifikation 4 %, nach Deskriptoren 33 %
Hersteller/Anbieter	IZ Sozialwissenschaften
Kosten	abhängig vom Host
Host(s)	STN, DIMDI, GBI / CD-ROM: WISO III

Abbildung 9

Name der DB	Sozialmedizin SOMED
angebot. Form(en)	online ☒ CD-ROM ☐ Diskettendienst ☐
Fachgebiet(e)	gesamter Bereich der sozialen (nicht-klinischen) Medizin und deren Rand-gebiete: Arbeitsmed., Gesundheitswesen & -erziehung, Suchtkrankheiten, med. Statistik & Epidemiologie, Umweltmedizin
Zahl der DE insgesamt	über 277.000 Jährl. Zuwachs: 15 - 20.000
abgedeckter Zeitraum	seit 1978 Update: monatlich
dokumentierte Materialien	Zeitschriftenaufsätze, Monographien, Sammelwerke und -beiträge, Dissertationen, Graue Literatur
Abstracts	ca. 62 % der DE
Sprachanteile	je 45 % Deutsch bzw. Englisch, Rest Französisch u. a.
für die EWI relevant	nach meinen Berechnungen ca. 25 % (Gesundheitserziehung, Sonderpädagogik, Suchtproblematik)
Hersteller/Anbieter	idis, Bielefeld
Kosten	keine Lizenzgebühr, Download 0,60 DM/Dok.
Host(s)	DIMDI

Abbildung 10

Name der DB	BRITISH EDUCATION INDEX
angebot. Form(en)	online ☒ CD-ROM ☐ Diskettendienst ☐
Fachgebiet(e)	Erziehungswissenschaft/Pädagogik Didaktik
Zahl der DE insgesamt	über 55.000 (BEI) & 10.300 (BETI)
abgedeckter Zeitraum	ab 1972
dokumentierte Materialien	Zeitschriftenaufsätze aus 200 - 250 päd. Zeitschriften (BEI), Dissertationen (BETI)
Abstracts	keine
Sprachanteile	ausschl. englischsprachige Lit.
für die EWI relevant	ja, 100 %
Hersteller/Anbieter	University of Leeds
Kosten	30 $/h & .30 $ pro Dokument
Host(s)	DIALOG demnächst auf der CD-ROM International ERIC

Abbildung 11

Name der DB	ERIC
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☐
Fachgebiet(e)	Pädagogik/Erziehungswissenschaft und Randgebiete
Zahl der DE insgesamt	ca. 700.000 jährlicher Zuwachs: ca. 31.000
abgedeckter Zeitraum	seit 1966 Update: monatlich
dokumentierte Materialien	Zeitschriftenaufsätze aus mehr als 700 vorwiegend amerikanischen päd. Zeitschriften, Forschungsreports
Abstracts	ja
Sprachanteile	ausschließlich englischsprachig
für die EWI relevant	ja, 100 %
Hersteller/Anbieter	ERIC
Kosten	60 $/h & .40 $ pro Dokument
Host(s)	DIALOG auszugsweise auf 2 CD-ROM

Abbildung 12

Name der DB	FRANCIS Sciences de L'éducation (520)
angebot. Form(en)	online ☒, CD-ROM ☐ Diskettendienst ☐
Fachgebiet(e)	Pädagogik/Erziehungswissenschaft, insbes. Geschichte & Philosophie der Erziehung, Bildungswesen, -politik, -ökonomie, -soziologie, Pädagogische Psychologie, Erwachsenenbildung u. a.
Zahl der DE insgesamt	über 110.000 Jährl. Zuwachs: 3.500
abgedeckter Zeitraum	seit 1972
dokumentierte Materialien	91 % Zeitschriftenaufsätze
Abstracts	83 % mit Abstracts
Sprachanteile	vorwiegend französischsprachig
für die EWI relevant	ja, 100 %
Hersteller/Anbieter	INIST
Kosten	k. A.
Host(s)	QUESTEL

Abbildung 13

Name der DB	PsycINFO
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☐
Fachgebiet(e)	Gesamte Psychologie, zudem relevante Gebiete der Psychiatrie, Soziologie, Erziehungswissenschaft, Medizin u. a.
Zahl der DE insgesamt	über 909.000 jährlicher Zuwachs: ca.30.000
abgedeckter Zeitraum	seit 1969 Update: monatlich
dokumentierte Materialien	Aufsätze aus ca. 1.500 Zeitschriften, Monographien, Dissertationen
Abstracts	77% mit Abstracts
Sprachanteile	vorwiegend Englisch
für die EWI relevant	geschätzt ca. 20 %
Hersteller/Anbieter	American Psychological Assoc.
Kosten	hostabhängig
Host(s)	DIMDI, DIALOG CD-ROM: PsycINFO

Abbildung 14

Name der DB	SocialSciSearch
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☐
Fachgebiet(e)	alle Gebiete der Sozial- und Geistes-wissenschaften
Zahl der DE insgesamt	über 2,5 Millionen Literaturhinweise jährl. Zuwachs: ca. 130.000
abgedeckter Zeitraum	seit 1973 Update: wöchentlich
dokumentierte Materialien	Aufsätze aus 1.500 Zeitschriften in Cover-to-cover-Auswertung und aus weiteren 2.400 Zeitschriften (Teilauswertung)
Abstracts	keine
Sprachanteile	vorwiegend englischsprachig, deutscher Anteil knapp 5 % (über 118.000 DE)
für die EWI relevant	ja, wenn englischsprachige Lit. gesucht wird.
Hersteller/Anbieter	Institute for Scientific Information (ISI)
Kosten	hostabhängig
Host(s)	DIMDI, DIALOG Auszug auf CD-ROM

Abbildung 15

Name der DB	Sociological Abstracts (SA)
angebot. Form(en)	online ☒ CD-ROM ☒ Diskettendienst ☐
Fachgebiet(e)	Soziologie, Sozialpsychologie, Anthropologie, Wirtschaftswissenschaften, Erziehungswissenschaft, Statistik u. a.
Zahl der DE insgesamt	über 385.000 jährl. Zuwachs: ca. 19.000
abgedeckter Zeitraum	ab 1963 Update: 5 x im Jahr
dokumentierte Materialien	Aufsätze aus mehr als 1.600 Zeitschriften; Monographien, Dissertationen, Konferenzberichte, Rezensionen
Abstracts	54 % mit Abstracts
Sprachanteile	80 % engl., 6 % franz., 3,5 % dt.
für die EWI relevant	ca. 5 %
Hersteller/Anbieter	Sociological Abstracts, Inc.
Kosten	hostabhängig
Host(s)	DIMDI, DIALOG CD-ROM: SOCIOFILE

Abbildung 16a: Hosts und deren Datenbanken in Auswahl

Host	Datenbanken	Bemerkungen
DIMDI = Deutsches Institut für medizinische Dokumentation und Information, Köln	CURRENT CONTENTS/SSCI, FORIS, (Gerontologie-DB GEROLIT), PsycINFO, PSYNDEX, PSYTKOM, SocialSciSearch, SOCIOLOGICAL ABSTRACTS, SOLIS, SOMED, (Sport-DB SPOFOR & SPOLIT) u. a.	Suchsprache: grips Zugang über öffentl. Telefonnetz, Datex-P, Datex-J/BTX, WIN, EUROPANET, INTERNET Über DIMDI kann man zum Host ECHO und dem DBI (ZBD1) wechseln. Mondscheintarif
STN International c/o FIZ Karlsruhe	BIBLIODATA, (Dissertations-DB DISSABS), FORIS, MATHDI, SOLIS, (Graue Literatur-DB SIGLE), (Umwelt-DB UFORDAT), (Umwelt-DB ULIDAT), (Lehr- & Forschungsst.-DB VADEMECUM)u. a.	Suchsprache: MESSENGER Zugang über öffentl. Telefonnetz, Datex-P, WIN, INTERNET Akademisches Programm (80% Rabatt)

Abbildung 16b

Host	Datenbanken	Bemerkungen
DIALOG europe, Frankfurt/M.	(Gerontologie-DB AGELINE), (audiovis. Material A-V-Online), BRITISH EDUCATION INDEX, ERIC, PsycINFO SocialSCISEARCH, SOCIOLOGICAL ABSTRACTS u. a.	Suchsprache: DIALOG Zugang über öffentl. Telefonnetz, Datex-P, Datex-J/BTX, WIN, INTERNET vorrangig englisch-sprachige DB
GBI = Gesellschaft f. betriebswirtschaftliche Information, München	FORIS, (Arbeitswissenschaft LABOR), PSYNDEX, SOLIS u. a.	Suchsprache: AOS
DBI = Dt. Bibliotheksinstitut, Berlin	BIBLIODATA, ZDB1, (Kongreßschriften-DB GKS1), (Nieders. Zs.-DB NZN0), (Bayr. Stabi-ZS-DB ZBSB), (Verbundkatalog MONO)	Suchsprache: grips keine Aufsätze in den DB, doch Online ordering von Aufsätzen über ZS.-Titel möglich
QUESTEL	FRANCIS u. a.	Suchsprache: QUESTEL

Alexander Botte

Die "Literaturdokumentation Bildung auf CD-ROM" des FIS Bildung

Eine erste Bilanz nach acht Monaten

Sehr geehrte Damen und Herren, wie viele von Ihnen sicherlich registrieren werden, erhalte ich hier auf der 2. Fachtagung der GIB zum zweiten Male die Gelegenheit, die Arbeit des Fachinformationssystems Bildung vorzustellen. Dafür möchte ich zunächst meinen Dank an die Initiatoren dieser Veranstaltung richten. Dies ist sicher einerseits Ausdruck der engen Verbundenheit der GIB mit unserem Modellversuch; ich hoffe aber andererseits, annehmen zu dürfen, daß unser Projekt "Literaturdokumentation Bildung auf CD-ROM" tatsächlich für den durch diese Tagung angesprochenen Personenkreis von besonderer Bedeutung ist und daher diese etwas herausgehobene Aufmerksamkeit zu Recht in Anspruch nimmt.

Wir haben mit der Herstellung dieser CD-ROM, d.h. konkret mit der Zusammenstellung einer aus relativ zahlreichen Zulieferungen gespeisten Literaturdokumentation Neuland betreten. Es gäbe verschiedene Aspekte dieses Unternehmens, die man einer näheren Betrachtungen unterziehen könnte. Frau Bambey wird dies morgen im Rahmen des Workshops 1 für unsere produktionsseitigen Erfahrungen tun. Ich möchte im Rahmen einer ersten Bilanz nach acht Monaten zu folgenden Fragestellungen berichten:

1. Akzeptanz der CD-ROM bei der anvisierten Klientel:
   - Anvisierter Abnehmerkreis,
   - Vertriebsbilanz,
   - Vertriebsperspektiven.

2. Perspektivische Verbesserungen der CD-ROM und Anpassung an die Nutzerwünsche:
   - Abdeckung der wesentlichen Inhaltsbereiche,
   - Weiterentwicklung des Rechercheinstrumentariums,
   - Steigerung der Aktualität der Information

3. Neuerungen auf der Ausgabe 1995 der CD-ROM:
   - quantitative Veränderungen,
   - Neuerungen im Bereich der Oberfläche und Präsentation.

1 Welche Akzeptanz hat die CD-ROM bislang bei der anvisierten Klientel gefunden und wie ist diese Akzeptanz einzuschätzen?

Wenn man die Marktaussichten eines Informationsproduktes, wie es die Literaturdokumentation Bildung auf CD-ROM darstellt, einschätzen will, so kann man sich wohl am ehesten am Vertrieb einer gedruckten Spezialbibliographie orientieren. Im Vergleich zum Druckwerk hat die CD-ROM die Vorteile, daß sie weit größere Anwendungsmöglichkeiten bietet und als relativ neuestes Informationsmedium besondere Attraktivität genießt. Andererseits ist sie gegenüber dem Druckmedium dadurch benachteiligt, daß sie nur für einen eingeschränkten Benutzerkreis überhaupt in Frage kommt, der nämlich über die technischen Voraussetzungen verfügt. Nachteilig wirkt sich sicher auch aus, daß es gerade im geisteswissenschaftlichen Bereich immer noch Vorbehalte, wenn auch abnehmende, gegenüber den elektronischen Informationsmitteln gibt.

Unter den genannten Voraussetzungen muß man den Kreis der Einrichtungen, die mit unserem Produkt CD-ROM realistischerweise erreicht werden sollten, in folgenden drei Bereichen sehen:

1. mittlere und größere wissenschaftliche Bibliotheken von Einrichtungen mit einem pädagogischen Schwerpunkt (u. a.: Hochschulbibliotheken, Fachbereichsbibliotheken, landesweite Einrichtungen der Aus- und Fortbildung, Forschungsbibliotheken),
2. Informationsvermittlungsstellen im pädagogischen Feld,
3. Zentrale Bibliotheken der Bildungsverwaltung und Bildungspolitik (Ministerien, Behörden).

Alle übrigen Abnehmerkreise, die man als potentielle bezeichnen könnte, haben unter gegenwärtigen Bedingungen nur eine eher eingeschränkte Nutzungserwartung an eine CD-ROM-Datenbank, die vor allem in der voraussichtlich zu geringen Häufigkeit der Inanspruchnahme ihre Begrenzung findet. Zu nennen wären hier:

- Einzelpersonen (Wissenschaftler, Studenten, Lehrer)
- Einrichtungen der pädagogischen Praxis (Schulen, Erwachsenenbildung usw.)

Soweit man die Verbreitung der CD-ROM im letzteren Feld der pädagogischen Praxis fördern will, muß man dies auch durch eine besondere Preisgestaltung unterstützen, die der geringeren Nutzungsfrequenz und auch der geringeren Kaufkraft dieser Klientel angepaßt ist. Wir haben dies bisher für den gesamten Bereich der Schule und der praktischen Lehreraus- und -fortbildung gemacht, indem wir diesen Einrichtungen auf Beschluß des Koordinierungsausschusses und auf Basis einer Vereinbarung mit der Softwarefirma eine besonders günstige sogenannte "Schulversion" zu 100.- DM anbieten können.

Wie sieht nun der Verkaufsabsatz der CD-ROM nach acht Monaten aus?

Bis Anfang November 1994 wurden als Einmalexemplar oder im Abonnement für Netz- und Einzelplatzversionen 116 Exemplare verkauft. Diese verteilen sich wie folgt auf spezifische Kundenkreise:

Art der Institution	*Verkauf*	*Freiexemplare*
Universitätsbibliotheken	31	
Fachbereichsbibliotheken (i.d.R. erziehungswiss. FB)	16	
Pädagogische Hochschulbibliotheken	13	
Sonstige Fachhochschulbibliotheken	8	
Landesinstitute, Landesmedienstellen, Landesbibliotheken, Ministerien	25	100
Sonstige Aus- und Fortbildungseinrichtungen	1	
Fachinformationseinrichtungen	8	17
Forschungseinrichtungen	5	
Stadtbibliotheken	3	
Verlage	2	
Privatpersonen	4	

Stand: 10.11.1994

Darunter befinden sich 20 Exemplare zum reduzierten Preis für die Schulversion. Wie man sieht, gehört dazu keine einzige Schule, sondern es handelt sich durchweg um Lehrerbildungseinrichtungen. Außerdem konnten die Herausgeber-Institutionen zu diesem Preis Zweitexemplare bestellen.

Diese Bilanz der Verkaufszahlen ergibt allerdings nur ein sehr unvollständiges Bild über den gesamten Vertrieb bzw. über die aktuelle Verbreitung der "Literaturdokumentation Bildung auf CD-ROM". Über den Verkauf hinaus wurden noch 117 Freiexemplare verschickt. Freiexemplare wurden selbstver-

ständlich vergeben an die 14 zuliefernden Institutionen, die dieses Projekt zusammen mit FIS Bildung als Herausgeber tragen. Einige CD-ROMs wurden als Tausch-Exemplare eingesetzt.

Weiterhin wurden auf Beschluß des Koordinierungsausschuß insgesamt 100 Freiexemplare nach einem bestimmten Proporz an die Kultusministerien der Länder vergeben, die diese im eigenen Haus und an Einrichtungen der pädagogischen Praxis weiterleiten konnten. Danach dürfen wir wohl davon ausgehen, daß zur Zeit alle Kultusbehörden und alle kulturhoheitlichen Einrichtungen der Lehreraus- und -fortbildung - soweit sie über die technischen Voraussetzungen verfügen! - mit unserer CD-ROM ausgestattet sind. Praktisch vollständige Verbreitung gilt wohl auch für die relativ überschaubare Zahl der freien pädagogischen Forschungseinrichtungen und der pädagogischen Informationsvermittlungsstellen, die ja großenteils identisch sind.

In Bezug auf die staatlichen Stellen und die Einrichtungen der Lehreraus- und -fortbildung können wir mit gewissen Abstrichen auch für Österreich und die Schweiz die Aussage treffen, daß eine weitgehende Verbreitung der CD-ROM gesichert ist. Allerdings sind in der Schweiz wie zum Teil auch in Österreich die Universitätsbibliotheken noch unterrepräsentiert. Zur Zeit befinden sich in Österreich 11 und in der Schweiz 7 CD-ROMs. Hinzu kommt im Ausland lediglich noch eine, die von einem pädagogischen Fachbereich einer belgischen Universität erworben wurde.

Bilanzierende Bemerkung

Betrachtet man nun vor dem Hintergrund dieser Zahlen und der Ausgangserwartungen die bisherige Vertriebsbilanz unserer CD-ROM, so kann man durchaus von einer Art "Erfolgsstory" sprechen.

Schon nach acht Monaten Vertriebszeit kann festgestellt werden, daß dieses Informationsmittel mit knapp 240 Exemplaren in größtenteils einschlägigen Multiplikatoreinrichtungen verankert ist. Dies wird auch deutlich im Vergleich mit den CD-ROM-Produkten der verwandten Fachinformationseinrichtungen der Sozialwissenschaftler und Psychologen, die vergleichbare Abonnentenzahlen erst nach zwei bis vier Jahren Marktpräsenz erreicht haben.

Ausblick: Was können wir zur weiteren Verbreitung der CD-ROM noch tun?

Zur besseren Ausschöpfung der Absatzmöglichkeiten bieten sich vor allem die Hochschulbibliotheken an. Dort, wo UBs einen Zugang zu CD-ROM-Laufwerken eröffnen, kommt es in der Regel zu hohen Auslastungen der Geräte. Der Multiplikatoreffekt ist hier besonders groß, die Hochschulbibliotheken sind die klassischen Einrichtungen der CD-ROM-Nutzung.

Wir können davon ausgehen, daß wir zur Zeit in der Bundesrepublik noch ca. 30 bis 40 Bibliotheken von Hochschulen, in denen Pädagogik eine relevante Rolle als Studienfach spielt, zusätzlich versorgen können. Wir sind optimistisch, daß wir im Laufe des nächsten Jahres, vor allem mit der neuen Ausgabe, weitgehend dieses Ziel erreichen können.

Außerhalb der Bundesrepublik sehen wir die Situation so: In Österreich und in der Schweiz erwarten wir bei Hochschulen noch einen zusätzlichen Absatz von ca. 10 Scheiben. Im übrigen Ausland werden die Absatzchancen dieser CD-ROM generell aufgrund des Schwerpunktes auf der deutschsprachigen Literatur nicht zu hoch einzuschätzen sein. Auch wenn wir die nächste Ausgabe mit einer englischsprachigen Oberfläche versehen, wird man in der Regel nur mit Deutschkenntnissen guten Gebrauch von der CD-ROM machen können.

Wie können wir zukünftig finanzielle Aspekte in unsere Vertriebsstrategie einbeziehen?

Eine Senkung des Preises wird im gesamten Hochschulbereich kaum zu einer nennenswerten Absatzsteigerung führen, da unsere CD-ROM preislich am unteren Rand vergleichbarer Produkte rangiert.

Dies heißt jedoch nicht, daß mit Blick auf die Hochschulbibliotheken die Finanzierungsfrage keine Rolle spielt. Die jüngsten Restriktionen in den Bibliothekshaushalten sind ja allgemein bekannt, und sie wirken sich vor allem negativ auf die Bereitschaft aus, feste, über Jahre verpflichtende Abonnements einzugehen.

Die Beschaffungspolitik der wissenschaftlichen Bibliotheken wird angesichts des ständig wachsenden Angebots an CD-ROM-Datenbanken selektiver werden. Kriterien der Auswahl werden dabei vor allem sein, wie die fächerübergreifenden Bibliotheken ihre defizitären Informationsbereiche sinnvoll ergänzen können und wie sie dies gleichzeitig möglichst rationell tun können.

Unter diesem Gesichtspunkt wird es gerade für Hochschulbibliotheken, die ja in der Regel ein sehr breites Fächerspektrum vertreten, interessant, mit einer CD-ROM mehrere verwandte Fachgebiete abzudecken. Nur geringe Erfolgsaussichten sehe ich in Zukunft für CD-ROMs, die Datenbanken präsentieren, die mehrfach angeboten werden.

Grundsätzlich sei die perspektivische Empfehlung ausgesprochen, die Angebote auf dem Medium CD-ROM gut zu koordinieren und ggf. auch eine Kooperation vorzuziehen, denn die Verwaltung einer stets größer werdenden Anzahl von CD-ROMs bedeutet zum Beispiel für die Hochschulbibliotheken einen größeren Belastungsfaktor bei Schulungsaufwand u.a.

Eine andere Belastung, die besonders leicht die großen Bibliotheken trifft, ist die Preispolitik der Softwarehäuser, die für Netzwerknutzung pauschal sehr viel höhere Lizenzgebühren verlangen als für Einzelversionen. Da eine spezifische CD-ROM in der Praxis nur sehr selten simultan von zwei oder gar 20 Terminals aus angesprochen wird, zielt diese Regelung auf eine potentielle Situation, die die absolute Ausnahme darstellt. Hier muß auf allgemeiner Ebene etwas geschehen. Die Bibliotheken selbst engagieren sich stark auf diesem Feld, und wir haben auf dem letzten Dataware-User-Treffen den Eindruck gewonnen, daß diese Firma da auch zu Konzessionen bereit ist.

CD-ROM-Vertrieb in Einrichtungen der pädagogischen Praxis

Nun noch einige Bemerkungen zum Ziel der Verbreitung der CD-ROM in Bereichen der Bildungspraxis, speziell in Schulen. Dieses Ziel war u. a. sicher auch als bildungspolitische Absicht mit der Einrichtung unseres Modellversuchs verbunden. Deshalb sollten wir neben dem besonderen finanziellen Angebot auch weitere Anstrengungen in dieser Richtung unternehmen.

Dabei muß man sich ehrlicherweise aber bewußt sein, daß die Realität des schulischen Alltags zur Zeit z. B. nur eher geringe Verwendungsmöglichkeiten für eine Literaturdatenbank auf CD-ROM bereithält. Sei es Schulverwaltung oder Unterrichtsvorbereitung, nur relativ wenige Situationen sind vorstellbar, in der tatsächlich eine Literaturdokumentation zur Unterstützung herangezogen würde.

Problematisch wird die Nutzung an Schulen vor allem auch durch die derzeitige Unmöglichkeit, eine rasche Literaturversorgung mit dem Literaturnachweis zu verbinden.

Ein zukünftig eventuell in der Bedeutung anwachsendes Feld der schulischen Nutzung ist die propädeutische Behandlung elektronischer Informationsmittel in der Oberstufe: In entsprechend ausgestatteten Schulen können Schülergruppen die bibliographische Recherche exemplarisch anhand unserer CD-

ROM einüben und dabei inhaltliche Impulse für bestimmte Schulfächer aufnehmen.

Zweifellos bedarf die Verbreitung solcher Nutzungsweisen noch der Zeit; vorläufig hängt sie sehr stark vom Engagement einzelner Personen in den Schulen ab. Langfristig können auf diesem Feld nur über die Vermittlung der Kultusverwaltungen und der Lehrerbildungseinrichtungen Erfolge erzielt werden.

2 Mit welchen Maßnahmen kann eine perspektivische Verbesserung der CD-ROM in Anpassung an die Nutzerwünsche herbeigeführt werden?

Zur Bewertung der Qualität einer CD-ROM-Literaturdatenbank können jenseits der Fragen der formalen Organisation im wesentlichen drei Leistungsbereiche herangezogen werden:

1. Vollständigkeit in Bezug auf das thematische Zielgebiet: Abdeckung der wesentlichen Inhaltsfelder,
2. Exaktheit und Tiefe der dokumentarischen Erschließung: Weiterentwicklung des Rechercheinstrumentariums,
3. Steigerung der Aktualität der Information: Verkürzung des Update-Intervalls.

Das erste Kriterium ist mit Sicherheit am schwierigsten zu fassen und empirisch zu überprüfen, wenn es um ein so komplexes Themengebiet wie das gesamte Bildungswesen geht. Außerdem ist "Vollständigkeit" ein für Dokumentare verhängnisvoller Begriff. So wird in der Regel jede Dokumentationsstelle scheitern, wenn sie nur ein oder zwei einschlägige Wissenschaftler die Datenbank befragen läßt. Mindestens einer von beiden wird das Fehlen eines eigenen Beitrags in einem just neu begründeten Mitteilungsblatt feststellen.

Empirisch fundieren läßt sich, daß unsere CD-ROM zu allen größeren Themen des Bildungsbereichs mit umfangreichen Literaturhinweisen aufwarten kann; dies gilt in gewissem Maße sogar auch für die erkannten sogenannten Defizitbereiche, d. h. für die Themenfelder, für die wir bislang keine zentral zuständige Zulieferinstitution haben. Drei dieser Defizitbereiche seien hier speziell betrachtet: Gesamtrecherchen im Themenfeld "Berufliche Bildung" erbringen über 22.000 Titel, im Themenfeld "Hochschule" über 11.000 und im Themenfeld "Migrations- und Ausländerpädagogik" über 5.000 Titel.

Eine qualitative Bewertung dieser Nachweise ist natürlich ohne ausführliche Expertise unmöglich. Dennoch scheint mir die Aussage erlaubt, daß aufgrund

der Erfahrung und des Renommees der an der CD-ROM beteiligten IuD-Stellen und der Vielseitigkeit des Herangehens an die Literatur mit hoher Wahrscheinlichkeit davon ausgegangen werden kann, daß auch für die Randbereiche die zentralen Literaturtitel erfaßt wurden.

Als vorläufiges Fazit zum Punkt inhaltliche Abdeckung der pädagogischen Themenbereiche möchte ich formulieren: Die relative Vollständigkeit der thematischen Abdeckung ist der entscheidende Vorzug dieser Literaturdokumentation, der ja letztlich auch mit dem hohem Aufwand der komplizierten Datenaggregierung erwirkt wurde und wird.

Verbesserungen durch neue Zulieferer

Verbesserungen sind jedoch auch in dieser Hinsicht möglich durch die Einbeziehung weiterer Zulieferer. Wir haben schon in der zu Beginn des nächsten Jahres erscheinenden Ausgabe für das Themengebiet "Hochschule" die Daten des Instituts für Arbeitsmarkt- und Berufsforschung hinzugewonnen.

Mit einer ersten, noch relativ kleinen Zulieferung ist die "Projektgruppe Hochschulforschung" an der nächsten Ausgabe beteiligt, von der wir ab 1995 für ihr Spezialgebiet zentrale und qualitativ hochwertige Erschließungen erwarten können. Mit Hilfe dieser neuen Partner kommen wir sicher auch im Themenfeld "Hochschulbildung" zu einer kompetenten Vertretung der relevanten Literatur.

Mit der nächsten Ausgabe wird auch die Bibliothek für Bildungsgeschichtliche Forschung mit ihrem einmaligen Bestand an historischer Bildungsliteratur auf der CD-ROM vertreten sein. Die Bibliothek ist der Fernleihe angeschlossen und unterstützt mit besonderen Serviceleistungen vor allem auch die Funktion der Literaturversorgung.

Weiterhin gehen wir davon aus, daß wir auf der übernächsten Ausgabe der CD-ROM die Integration der Daten der "Literaturdokumentation Berufsbildung" in der einen oder anderen Form auf der CD-ROM vollziehen können. Diese Datenbank wird bis Frühjahr nächsten Jahres durch ein Projekt vorbereitet und dann am BIBB federführend weiterbetreut. Damit dürfte dann die CD-ROM auch auf diesem Gebiet nicht nur mit ausreichender, sondern mit besonderer Präsenz aufwarten können.

Die Einbeziehung der Schweizer Daten ist für die Ausgabe 1995 nicht möglich gewesen. Dieses Vorhaben muß auf die Ausgabe 1996 verschoben werden, weil in der Schweiz zur Zeit eine Neuorganisation und Zentralisierung der Literaturdokumentation im Gange ist, die die dort vorhandenen relativ geringen Ressourcen noch bindet. Wir erhoffen uns jedoch von diesem Schritt

noch einmal einen besonderen Impuls für den Absatz der CD-ROM in Schweizer Hochschulen, der bisher noch eher gering ist.

Mit den soweit geplanten Erweiterungen des Zulieferkreises dürften wir bis spätestens 1996 ein von der Integrationsleistung her so interessanten Informationsdienst anbieten, daß eine im Sinne der oben angesprochenen strengeren Selektion bei der Anschaffung von CD-ROMs durch Hochschulbibliotheken ausreichende Attraktivität gewährleistet sein sollte, um der CD-ROM einen dauerhaften Platz in diesem Abnehmerkreis zu sichern.

Die Strategie, durch quantitative Ausdehnung der Zahl der Zulieferer an FIS Bildung die Qualität der CD-ROM zu verbessern, sollte damit aber zu einem gewissen Abschluß kommen. Wir werden dann insgesamt 19 Datenzulieferungen in der Geschäftsstelle zu integrieren haben, was auch bei bester Routinisierung der Abläufe einen ganz erheblichen Zeitaufwand bei jedem Update unvermeidbar macht.

Leicht nachvollziehbar ist, daß sich die quantitative Erweiterung deutlich mit den anderen Zielen der qualitativen Verbesserung, wie der Weiterentwicklung der dokumentarischen Erschließungsleistungen und der Steigerung der Aktualität unserer Daten stößt. Verbesserungen auf diesen beiden Feldern sind unter den bei uns gegebenen heterogenen und komplizierten Produktionsbedingungen besonders schwer zu erreichen. Gleichzeitig sind sie aber von hoher Bedeutung für die Zufriedenheit mit dem Produkt, und sie sind gerade dann von steigender Dringlichkeit, wenn die Datenbank weiter anwächst.

Welche Möglichkeiten bestehen zur Verbesserung der Erschließungsleistungen?

Das zentrale Instrument der Recherche sind die Schlagwörter bzw. Deskriptoren. Angesichts der dezentralen Herstellungsweise der Daten und angesichts des gewaltigen Bestandes an historischen Dokumenten können wir keine Perspektive ins Auge fassen, die von der Zugrundelegung eines konsistenten, für alle Dokumente gültigen Deskriptorenbestandes ausgeht. Wir sind darauf angewiesen, die vorgefundenen Schlagwörter für die Erschließung zu verwenden und den einzelnen Kooperationspartnern ihre Freiheit auch zukünftig zu belassen, ihre themenspezifischen Varianten in ihrer Verschlagwortung zum Ausdruck zu bringen.

Daraus folgt, daß wir bezüglich der Schlagwörter den eingeschlagenen Weg der Synonymkontrolle weiterverfolgen und optimieren müssen. Lediglich im Bereich der formalen Ansetzung (Verwendung von nachgestellten Kommata z. B.) werden wir uns um die Festlegung von Regeln bemühen.

Bei einer Datenbank unserer Größenordnung wäre die Ergänzung der Deskriptoren-Abfrage durch die Möglichkeit einer systematischen Recherche aufgrund eines Klassifikationschemas wünschenswert. Wir haben auf der letzten Tagung der Zulieferer beschlossen, ab 1995 eine neue Systematik anzuwenden. Allerdings ist die Übernahme der Systematik für die Kooperationspartner freiwillig; sie ist lediglich verbindlich für die Zulieferungen zur *Bibliographie Pädagogik*. Obwohl die Verwendung dieser Systematik auf der CD-ROM gravierende Probleme aufwirft, wäre es perspektivisch bei einem erwarteten jährlichen Zuwachs von 25.000 bis 30.000 Dokumenten durchaus sinnvoll, baldmöglichst eine parallele systematische Strukturierung der Daten einzuführen. Wir werden dies auf der nächsten Tagung noch einmal zum Gegenstand machen.

Verbesserung der Aktualität der Daten

Wie schon erwähnt, beschleunigt die große Zahl der Zulieferer und die Tatsache, daß vorläufig noch mit jedem Update völlig neue Datenstrukturen zu integrieren sind, nicht gerade unsere Herstellungsabläufe in der Geschäftsstelle. Bei Konsolidierung des Zulieferstammes und bei Routinisierung der Abläufe kann perspektivisch wohl davon ausgegangen werden, daß zwischen dem Zeitpunkt der letzten Datenzulieferung und dem Erscheinen der CD-ROM nicht mehr als drei Monate liegen. Diese Zeitspanne wird aber kaum zu unterbieten sein.

Es wird angesichts unserer dezentralen Produktionsweise also auch zukünftig so sein, daß wir in puncto Aktualität hinter den Anbietern zurückstehen, die im wesentlichen in zentraler Weise die dokumentarische Erschließung betreiben.

Welche Möglichkeiten der Optimierung bleiben uns?

Sicher werden wir noch vor Ablauf des Modellversuchs im Jahre 1996 mit unseren Partnern überlegen, ob wir auf ein halbjährliches Update kommen könnten. Von der jährlich anfallenden Datenmenge her wäre ein solcher Aktualisierungsrhythmus mehr als gerechtfertigt. Fraglich ist, ob alle Beteiligten, auch die Geschäftsstelle von FIS Bildung, dabei verläßlich mitziehen können. Wir werden dies prüfen.

Eine weitere Möglichkeit wäre, den gedruckten Dienst, die *Bibliographie Pädagogik*, in neuer Weise mit der CD-ROM zu koppeln. Die Idee ist, mit einer eventuell stark verkleinerten Zahl der Zulieferer, die einen übers Jahr kontinuierlichen Datenzuwachs garantieren können, eine vierteljährliche Aktualisierung zwischen den CD-ROM-Ausgaben zu produzieren, die als ge-

druckter Dienst in einfacher Ausstattung erscheint. Die CD-ROM und die gesamte Bibliographie Pädagogik folgten weiterhin im jährlichen Abstand.

3 Welche Neuerungen bringt die neue Ausgabe der CD-ROM?

Quantitativer Datenzuwachs

Die nächste Ausgabe wird statt bisher 14 nun 17 zuliefernde Institutionen verzeichnen. Die Datensammlung wird deutlich anwachsen um ca. 30.000 neue Nachweise (47.000 neu zugelieferte Daten vor Duplizitätsprüfung). Die neuen Daten sind überwiegend tatsächlich neu, also aus den Erscheinungsjahren 1993 und 1994. Gleichzeitig haben wir eine intensive Nachkorrektur der alten Daten vorgenommen und mit einem neuen Duplizitätsprüfcode den alten Datenbestand auf Dubletten durchforstet.

Leicht gestiegen ist auch der Anteil an Referaten, der bei den neuen Erschließungen über 24 % liegen wird; gegenüber ca. 12 % im bisherigen Gesamtdatenbestand. Weiterhin besteht hier natürlich im Vergleich zu anderen Produkten dieser Art eine bedeutsame Diskrepanz.

Neuerungen im Bereich der Oberfläche

Wir haben eine ganze Reihe vor allem kleinerer Verbesserungen bei der Handhabung und der Oberflächenoptik der CD-ROM vorgenommen. Dabei wurde jedoch davon ausgegangen, daß sich am grundsätzlichen Erscheinungsbild möglichst wenig ändern sollte. Nur die wichtigsten Veränderungen seien hier erwähnt.

Es wird in der nächsten Ausgabe eine englische Sprachversion zur Verfügung stehen, was bedeutet, daß die Software-Oberfläche und die Feldbezeichner in dieser Sprache vorgefunden werden. Dabei hat es keine Veränderungen an den Daten gegeben; weder wurden die Deskriptoren noch etwa die Sachtitel ins englische übersetzt. Inwieweit dieses Angebot den Absatz im nichtdeutschsprachigen Ausland fördert, bleibt abzuwarten.

Ab sofort kann auch die lange angekündigte Macintosh-Plattform eingesetzt werden, allerdings bisher nur in englischer Sprache. Da die deutschsprachige Oberfläche und die Hilfstexte für die allernächste Zeit erwartet werden, steht sie wahrscheinlich mit dem Erscheinen der neuen Ausgabe zur Verfügung.

Eine wichtige Neuerung betrifft die Einführung einer produktspezifischen Hilfefunktion für die Kategorien der Suchformate. Für jede Suchkategorie kann mit F1 eine von uns verfaßte Hilfe aufgerufen werden. Für die Ausgabe der Dokumente sind solche produktspezifischen Hilfefunktionen nicht möglich, so daß wir leider bestimmte Felder, die nur zur Ausgabe vorgesehen sind, über den Bildschirm nicht beschreiben können.

Präsentation

Bei den Anzeige- und Ausgabeformaten haben wir - entsprechend den zu uns gedrungenen Nutzerwünschen - einige Ergänzungen vorgenommen, soweit dies mit vertretbarem finanziellen Aufwand realisierbar war. So gibt es jetzt weitgehend die von anderen Datenbankanbietern gewohnten Standardformate auch bei uns.

Des weiteren haben wir einige zusätzliche Indexierungen in Auftrag gegeben. So können zukünftig Deskriptoren und Personennamen sowohl im Feldindex, wie bisher, als auch im Wortindex gesucht werden. Für die Felder "Titel" und "Abstract" werden auch die Umgebungsoperatoren indiziert. Das heißt, es können logische Und-Verknüpfungen auf bestimmte "Nahe"-Distanzen begrenzt werden.

Soweit zur Ausgabe 1995. Lassen Sie mich noch die folgende **Schlußbemerkung** anfügen. Im Rahmen des Modellversuchszeitraums, der ja auf die Jahre 1995/96 verlängert worden ist, werden wir eine empirisch gestützte Analyse von Bildungsinformationsdiensten vorlegen. Im Wege der Erstellung dieser Untersuchung werden wir auch eine Befragung der Nutzer unserer CD-ROM durchführen. Diese Befragung wird sicher weiteren Aufschluß geben, in welcher Richtung Veränderungen und Neuentwicklungen gewünscht werden. Wir sind aber auch im gegenwärtigen Stand schon überzeugt, daß das Konzept Literaturdokumentation Bildung auf CD-ROM noch einige Jahre seine Tragfähigkeit unter Beweis stellen wird.

Anschrift des Verfassers

Alexander Botte
Leiter Modellversuch FIS Bildung
Berliner Str. 31-35
65760 Eschborn
Tel.: (06169) 4841-45; Fax: -61

Rosemarie Kohls

Zusammenfassung der Diskussion zu den Vorträgen von Rost und Botte

Entsprechend dem Anliegen unserer Gesellschaft möchten wir auch diesmal wieder Produzenten, Vermittler und Nutzer von Informationen (einschließlich Dokumentationen) zusammenführen, um gemeinsam darüber zu beraten, welche Informationen im Bildungsbereich gebraucht werden und wo bzw. wie sie rationell erstellt und abgerufen werden können. Von großem aktuellen Interesse ist hierbei, welche neuen Möglichkeiten der modernen Informations- und Kommunikationstechniken dafür bereits heute genutzt werden können - wie unsere Tagung insgesamt zeigt.

Auch mit den Beiträgen der zwei erfahrenen GIB-Kollegen, Dr. Friedrich Rost und Alexander Botte, wollten wir aktuell informieren und Erfahrungen vermitteln. Zugleich erhalten alle Teilnehmer die Gelegenheit zu produktiver Diskussion und Anfragen. Dabei sollte der Finger auch auf die Stellen gelegt werden, die wir im Bildungsbereich weiter verbessern müssen.

Beitrag Rost

Zur Datenbank MATHI erfolgten seitens des Herstellers, Herrn König vom Fachinformtionszentrum Karlsruhe, kurze ergänzende Aussagen. U. a. werden vom Hersteller Auftragsrecherchen in hoher Qualität ausgeführt, Kosten: 200,-- bis 400,-- DM.

Hinsichtlich der Qualität bildungsrelevanter Datenbanken wurde mehrfach auf die notwendige weitere Verbesserung informationssprachlicher Fragen hingewiesen. Fachlich kompetente, inhaltliche Aufbereitung und exakter Umgang mit Fachbegriffen werden als notwendiger Anspruch gekennzeichnet.

Studierende sind noch besser als Nutzer zu berücksichtigen. Sie recherchieren in der Regel nicht online, sondern primär mit CD-ROMs. Hierfür ist der methodische Zugang durch entsprechende Handbücher und Unterweisungen

weiter zu verbessern. Für Auftragsrecherchen sind akzeptable Gebühren zu erheben (z. B. 30,-- DM/Rechercheauftrag).

Bei großen Datenhosts können Sonderkonditionen für Bildungszwecke genutzt werden (z. B. Akademisches Programm, Mondscheintarif bei STN).

Beitrag Botte

In der Diskussion gab es mehrere positive Bewertungen zur CD-ROM Bildung - gute praktikable Vermittlung und didaktische Aufbereitung. Die CD-ROM sollte aber noch breiter bekannt gemacht werden.

Folgende Vorschläge für weitere Verbesserungen sind unterbreitet worden:

- Künftig sollten auf der CD-ROM auch wichtige Beiträge aus Handbüchern, Enzyklopädien, Sammelwerken und die Neuauflagen verzeichnet werden.
- Die Nutzeroberfläche ist weiter zu verbessern, insbesondere aus studentischer Sicht. Dabei sind mehrere Betriebssysteme zu beachten.
- Ein methodisches Handbuch oder evtl. auch Beiheft für Studenten bzw. eine Produktbeschreibung sind zu erstellen mit Vergleichen zu anderen vorliegenden Angeboten. Hierzu könnte z. B. auch eine Diplomarbeit vergeben werden.
- Für die Netzwerk-Version und Privatnutzer (Wissenschaftler, Studenten) sind kostengünstigere Preisangebote zu entwickeln.
  Es wird auch empfohlen, das bisherige Verteilungsprinzip der CD-ROM Bildung (Freiexemplare) künftig zu überdenken und auf höchste Wirksamkeit zu achten - bis in die Schulen hinein.

In diesem Zusammenhang wurden auch Fragen einer hohen Aktualität der nachgewiesenen Informationsquellen angesprochen und darüber informiert, daß Herr Heiner Schmidt, der bisherige verdienstvolle Bearbeiter von "ZEUS", seine Dokumentationsarbeit einstellen wird. Wie kann danach eine aktuelle (gedruckte) Information gesichert werden? Hierzu haben bereits mehrere Gespräche stattgefunden. Diese Thematik soll recht bald weiter erörtert und zufriedenstellend gelöst werden.

Anschrift der Verfasserin
Dr. Rosemarie Kohls
Projektgruppe Hochschulforschung
Literaturdatenbank Hochschulbildung
10318 Berlin
Aristotelessteig 4
Tel.: (030) 50 19 26-78; Fax: -41

Rainer Dechsling

Versuch einer Typologie der textorientierten elektronischen Fachinformation

Datenbank, Hypertext oder linearer Text?

Die elektronische Fachinformation bestand bislang vorwiegend aus klassischen Datenbanken. Mittlerweile zeichnet sich hier jedoch ein Wandel ab, der vor allem durch CD-ROM-publizierende Fachverlage getragen wird. Neben das auf dem Online-Markt dominierende Modell der klassischen Datenbank treten neue Formen der elektronischen Textvermittlung. Im folgenden wird eine Typologie der textuellen Information versucht. Erwägungen zur korrespondierenden Abfragesoftware speziell im CD-ROM-Bereich und ein Ausblick auf die nähere Zukunft sollen den Anbietern elektronischer Fachinformation als Entscheidungshilfe dienen.

Textuelle Information tritt grundsätzlich in zwei vorwiegend nicht-linearen Formen (Datenbank, Hypertext) sowie in linearer Gestalt (Buch) auf:

1. **Klassische Datenbank:** Bei den existierenden Nachweis- und Volltextdatenbanken handelt es sich zumeist um nicht-linearen Text, der selektiv erschlossen und gelesen wird. Die einzelnen Nachweise, Beiträge bzw. Datensätze haben thematisch meist nur locker miteinander zu tun. Niemand käme auf die Idee, derartige Texte von vorn bis hinten "durchzulesen".
2. **Hypertext:** Viele Verlage und Institutionen sind im Begriff, ganz eigenständige elektronische Informationsprodukte zu entwickeln, bei denen nicht die Sammlung möglichst vieler Datensätze, sondern die gezielte Wissensvermittlung im Vordergrund steht. Dabei werden Texte ganz bewußt als kurze und in sich verständliche Einheiten konzipiert und sinnvoll miteinander verknüpft. Die Erstellung und bildschirmgerechte Darbietung der Texte ist nicht-linear. Es handelt sich um klassischen Hypertext, der heute meist mit Multimedia-Elementen (Bild, Ton, Video) angereichert wird.
3. **Elektronisches Buch:** Wohl nur für die Verleger klassisch-sequentieller Bücher stellt sich die Anforderung, den Lesekomfort sorgfältig gestalteter Bücher in das elektronische Medium zu überführen. Am PC sollen mindestens auch diejenigen Zugangswege zum Text eröffnet werden, die ein gut

strukturiertes Fachbuch bietet. (Mit dem "elektronischen Buch" sind nicht die "electronic books" eines des weltweit größten Medienkonzerne gemeint. Bei diesen handelt es sich häufig um Lexika und somit um nicht-linearen Text.)

Jede dieser drei Textgestalten findet ihre Entsprechung in spezifischer Abfragesoftware. Die allumfassende marktgängige Software, die alle drei Ausprägungen der Information vollständig abdeckt, gibt es derzeit nicht. Die nachfolgenden Ausführungen sollen Informationsanbietern und Fachverlagen helfen, den nach dem heutigen Stand der Technik richtigen Softwaretyp zu finden.

Nachweis- und Volltextdatenbanken

Der Kategorie der Nachweisdatenbanken dürfte derzeit im Bereich der elektronischen Fachinformation die größte wirtschaftliche Bedeutung zukommen. Kataloge und Verzeichnisse wie das *VLB, Wer liefert was, Perinorm* (mehrsprachiger Nachweis technischer Regeln) oder die Leitsatzkartei des deutschen Rechts sind CD-ROM-Projekte, die Märkte geöffnet haben. Der durchschlagende Erfolg dieser Nachweisdatenbanken beruht u. a. auf der weitverbreiteten und zutreffenden Erwartungshaltung der Käufer, daß Kataloge und Verzeichnisse zur Umsetzung auf den PC bestens geeignet seien. Aus der Sicht der Verlage lassen sich derartige Projekte relativ unproblematisch als Database Publishing organisieren. Die Substitution der zugrundeliegenden Druckwerke ist erheblich.

Meist erst im zweiten Schritt werden textorientierte Zeitschriften im Volltext zur Verfügung gestellt. Es handelt sich um Volltextdatenbanken, die als raum- und zeitsparendes elektronisches Archiv dienen. Eine nennenswerte Substitution der gedruckten Abonnements durch das elektronische Angebot findet bislang kaum statt, weil es am Umfeld (z. B. Anzeigen und Begleitinformationen im Umschlagteil) fehlt. (Für die aktuelle elektronische Zeitschrift, auf die hier nicht eingegangen werden kann, gelten ganz andere Mechanismen. Dafür bieten sich Tools wie Adobe Acrobat an.)

Die Abfragesoftware bei Nachweis- und Volltextdatenbanken sollte ungefähr die folgenden Charakteristika aufweisen:

- Abbildung einer differenzierten **Feldstruktur** in Suchmaske, Trefferliste und Vollanzeige. Möglichkeit des Informationsanbieters bzw. Verlages, den Anwendern mehrere Suchmasken, Trefferlisten und Vollanzeigen zur Auswahl anzubieten, um trotz komplizierter Feldstruktur eine einfache und transparente Recherche zu ermöglichen. Felder vom Anwender einzeln ein- und auszublenden.

- Feldbezogene **Suchoptionen**: Indizes jeweils mit einem Tastatur- oder Mausklick aufzuschlagen. Einfache Navigation im Index, z. B. Positionierung aufgrund Benutzereingabe. Markierung im Index. Direkte Übernahme von Suchbegriffen aus dem Index für die Suche. Synonymfunktion und Thesaurus. Suchoperatoren: UND, ODER, OHNE, NICHT, >, >=, <, <=, =, Bereich, Wort-, Satz- oder Absatzabstand. Flexible Datumssuche (Mai 1985, 5.85, 1.5.85, 1985-05-01 etc.). Schnelle Links-, Mitte- und Rechtstrunkierung. Benutzerkontrollierte Trunkierungsauflösung. Grundformenerkennung. Geschachtelte Klammersetzung, feldübergreifende Verknüpfung. Wahlweise unscharfe Suche (fuzzy logic). Automatische Rechtstrunkierung. Logisch (Ja/Nein). Numerisch. Numerisch in Kombination mit Wort, Zeile, Feld. Zugriff auf frühere Suchen.

- Weitere **Indexoptionen** des Informationsanbieters bzw. Verlages: Indizierung nach Wort, Zeile oder Feld. Sprache (Deutsch, Englisch, Amerikanisch, Französisch etc.). Stopwortliste. Differenzierte Behandlung von Interpunktion und Akronymen. Bindestrichbegriffe ein-, zwei- oder dreifach indizierbar. Optimierung der Suchgeschwindigkeit nach dem Zahlenverhältnis der unterschiedlichen Indexeinträge und der nachgewiesenen Datensätze. Gemeinsames Indizieren mehrerer Felder in beliebiger Kombination. Mehrere Indizes pro Feld.

- **Navigation**: Paralleles Suchen in mehreren Fenstern. Querverweissuche aus dem gefundenen Datensatz heraus (soft wired hypertext). Verfolgung fester Verweise (hard wired hypertext) mit einem Klick. Jederzeitiges Umschalten zwischen Suchmaske, Trefferliste und Vollanzeige sowie zwischen verschiedenen Suchen unter vollständiger Erhaltung des jeweiligen Status. Sortierung oder Ranking der gefundenen Datensätze nach verschiedenen Kriterien. Markierung einzelner Datensätze in positiver oder negativer Selektion für die Weiterverarbeitung (Anzeige, Ausdruck, Speicherung, Zwischenablage). Kopf- und Fußzeile in der Vollanzeige. Präzises Highlighting entsprechend der eingegebenen Suche. Präziser Direktsprung zu den gesuchten Passagen des Datensatzes. Expliziter Sprung zum nächsten oder vorigen gefundenen oder benachbart gespeicherten Datensatz (logisches und lineares bzw. physisches Blättern). Volle Zitierfähigkeit nach Seiten- oder Blattzahl.

- **Vollanzeige**: Zügiges Blättern auch in sehr langen Datensätzen (Volltext größer 64 KB). Unterstützung von Fließtext und geschützten Bereichen (Tabellen). Typographische Hervorhebungen auf Feldebene und innerhalb von Feldern (bei zeichenorientierten DOS-Anwendungen wenigstens Highlighting und Invertierung). Darstellung von Formeln und Graphiken. Zoomfunktion für Graphiken. Für den Informationsanbieter bzw. Verlag: Ausgefeilte Anzeigelogik auf Feldebene.

- **Suchgeschwindigkeit**: Im Sekundenbereich auch in CD-ROM-Indizes mit mehreren Millionen Einzelvorkommnissen. Dies auch bei Trunkierung und Abstandssuche. Meldung der Trefferzahl vor Aufbau der Trefferliste. Zügiger Aufbau der Trefferliste. Möglichkeit zur Vollanzeige vor Aufbau der Trefferliste.

- **Produktionsprozeß**: Unterstützung diverser Text-, Graphik- und Datenbankformate beim Einlesen der Daten (SGML, RTF, BMP, TIF, dBase, Oracle, ASCII oder ANSI mit frei definierbaren Delimitern etc.). Umsetzung von Zeichen und Zeichenketten. Schnelle Indizierung auch bei sehr großen Datenbeständen. Menügeführte Anwendungsentwicklung ohne Programmierung. Umfangreiche Statistiken und Testmöglichkeiten. Vollautomatische Erzeugung des Installationsprogramms und des Premasters für die CD-ROM.

Ohne Anspruch auf Vollständigkeit seien die in Deutschland marktgängigen Softwareprodukte genannt, welche die genannten Anforderungen weitgehend erfüllen: CD Answer / CD Author (Dataware Technologies), Cobra (EPS), OptiSearch (MPW Lasec), CDIS (Makrolog).

Grundsätzlich abzuraten ist von relationalen Datenbanken ohne Volltextkomponente (Clipper und andere dBase-Compiler, MS Access Distribution Kit etc.), denn die meisten Kataloge und Verzeichnisse enthalten bei genauerem Hinsehen Felder, in denen eine Volltextsuche bessere Ergebnisse liefert als die dBase-übliche indizierte Feld- bzw. nichtindizierte sequentielle Suche. Auch die nachfolgend unter "Hypertextsysteme" und "Buchkompatible Software" aufgeführten Softwareprodukte sind für Nachweis- und Volltextdatenbanken in der Regel ungeeignet, da ihnen wichtige Funktionalität wie z. B. differenzierte Suchoptionen und Trefferlisten fehlen.

Hypertextsysteme für die elektronische Wissensvermittlung

Informationsorientierte elektronische Medien leben jedoch nicht zwangsläufig von der Quantität der darin gespeicherten Daten. Bisweilen kommt es auch auf die Originalität der dargebotenen Information an. Es gibt beispielsweise im juristischen Bereich Verlage, die primär für das elektronische Medium nicht-linearen, verknüpften Text produzieren lassen. Anhand einer detailliert vorgegebenen Systematik werden zahlreiche Einzelthemen von externen, per Verlagsvertrag gebundenen Autoren bearbeitet. Jedes Thema muß aus sich heraus verständlich sein und darf nicht länger als zwei bis drei Bildschirmseiten (mit jeweils ca. zwanzig Zeilen) sein. Die elektronisch dezentral erfaßten Texte werden von einem fachlich versierten Herausgeber bzw. Wis-

sensingenieur überprüft und mit den übrigen Themen verknüpft. Multimediale Erweiterungen mit Bild, Ton und Video sind möglich.

Damit ist auch schon der klassische, auf nicht-linearem Text basierende Hypertext-Ansatz im Grundsatz beschrieben. Die Navigationsprobleme in einem derart dezentral vernetzten Textraum sind beträchtlich. Dennoch kann es sinnvoll sein, dem Ausruf "Lost in hyperspace!" zum Trotz nicht-lineare Informationssysteme zu schaffen. Besonders im institutionellen Bereich (Großunternehmen, Museen etc.) spielen Hypertextsysteme schon heute eine bedeutende Rolle. Auch im Verlagsbereich gibt es Projekte, bei denen ein originäres Hypertextsystem der schlichten Umsetzung vorhandener Druckwerke vorzuziehen ist. Dabei kann durchaus auf den bestehenden Autorenstamm zurückgegriffen werden, soweit er dem PC aufgeschlossen gegenübersteht.

Als typische Hypertextprodukte sind zu nennen: Hypercard (Apple), Toolbook, Knowledge Pro, Multimedia Viewer (Microsoft). Der Übergang zum Multimedia-Bereich ist fließend. Dort gibt es weitere Tools.

Buchkompatible Software für linearen Text

Bislang wurden die Nachweis- und Volltextdatenbanken sowie der klassische Hypertext als überwiegend nicht-lineare Textgattungen behandelt. Von der ersten bis zur letzten Seite linearer Text findet sich herkömmlich vor allem in Büchern, sei es als Shakespeares Hamlet oder als technisches Handbuch. Wenn nun von "buchkompatibler Software" gesprochen wird, ist damit also Software gemeint, die zur Darstellung von linearem Text besonders geeignet ist. Die Arbeitsgruppe "CD-ROM/informationsorientiert" des Arbeitskreises Elektronisches Publizieren im Börsenverein des Deutschen Buchhandels hat hierzu im März 1994 die "Leipziger Empfehlungen" verabschiedet.

Wichtige Anforderungen an buchkompatible Software sind:

- Buchtypographie am Bildschirm und im Ausdruck,
- Zugang auch über Gliederung (Inhaltsverzeichnis),
- sequentielle Darstellung des linearen Textes,
- komfortable Fußnotenverwaltung,
- Notizfunktion, Lesezeichen, Exzerpieren,
- lebender Kolumnentitel bzw. mitlaufende Gliederung,
- dreidimensionales Blättern (von Abschnitt zu Abschnitt außerhalb der Treffermenge - zum nächsten oder vorigen gefundenen Datensatz - Direktsprung von Treffer zu Treffer).

Für die oben beschriebenen nicht-linearen Informationssammlungen sind diese Anforderungen teilweise wenig bedeutsam oder sogar unpassend. So ist

etwa der Zugang über eine Gliederung bei Nachweis- und Volltextdatenbanken häufig nur um den Preis intensiver redaktioneller Nachbearbeitung realisierbar, weil die zugrundeliegenden Kataloge, Verzeichnisse und Fachzeitschriften lediglich alphabetisch oder chronologisch geordnet und bestenfalls mit wenig interessanten, ständig wiederkehrenden Rubriken versehen sind. Auch die sequentielle bzw. lineare Textdarstellung ist in nicht-linearen Datenbanken regelmäßig belanglos. Sinn macht darin allenfalls das lineare (bzw. sequentielle bzw. physische) Blättern außerhalb der Treffermenge, um die ursprüngliche alphabetische, chronologische oder sonstige Ordnung am Bildschirm nachvollziehen zu können. Auch das kann aber durch eine parallele Suche z. B. nach den Seiten 800 bis 850 eines bestimmten Zeitschriftenjahrgangs ersetzt werden, wenn nur die Datensätze in der Folge des gedruckten Vorbilds auf der CD-ROM (oder irgendeinem anderen Medium) gespeichert sind. Wenig sinnvoll wäre es, sämtliche Datensätze einer nicht-linearen Datenbank als Endlostext darzustellen, da das kaum merkliche Überfliegen von Datensatzgrenzen die konzentrierte Arbeit in heterogenen Datenbeständen erschwert.

Ergänzende Anforderungen an buchkompatible Software ergeben sich aus der bei Nachweis- und Volltextdatenbanken üblichen Funktionalität (s. oben). Einzelheiten können den bereits erwähnten "Leipziger Empfehlungen" entnommen werden.

Folgende in Deutschland gängige Softwareprodukte verdienen das Etikett "buchkompatibel": CD Answer Hypertext (Dataware Technologies), Folio Views, Dynatext. Mit der Einschränkung, daß der lineare Text am Bildschirm nicht sequentiell, sondern datensatzbezogen mit der Möglichkeit jederzeitigen linearen Blätterns dargestellt wird, sind auch CDIS (Makrolog) und Multimedia Viewer (Microsoft) zu nennen. Auch diese Nennungen können nicht den Anspruch auf Vollständigkeit erheben.

Konvergenz und Konnektivität

Eine für alle linearen und nicht-linearen Textgattungen gleichermaßen perfekt geeignete Abfragesoftware gibt es derzeit nicht. Der Versuch, das textorientierte elektronische Publizieren auf ein einziges Softwareprodukt zu vereinheitlichen, ist beim jetzigen Stand der Technik zum Scheitern verurteilt, weil kein Informationsanbieter oder Verlag seinen Kunden den Verzicht auf werkbezogen sinnvolle und längst selbstverständliche Funktionalität zumuten kann. Das ist indes kein Grund zur Resignation:

1. Die Anwender haben unter den immer stärker dominierenden graphischen Benutzeroberflächen (Windows, Apple etc.) viel weniger Aufwand, sich in neue Anwendungen hineinzufinden, als dies unter DOS der Fall ist (war).

Die Bedienerfreundlichkeit ist deshalb kein besonders starkes Argument für eine Festlegung der Informationsanbieter und Verlage auf eine bestimmte Abfragesoftware.

2. Die Nachfragemacht der Informationsseite gegenüber den Herstellern und Distributoren von Abfragesoftware ist gewiß dann am größten, wenn sie gebündelt auftritt. Andererseits brächte die konsequente Einigung auf ein bestimmtes Abfragetool auch einen Verlust an Konkurrenz, der die technische Entwicklung zu behindern geeignet wäre. Es fragt sich deshalb, ob es eine gute Strategie wäre, durch die gezielte Förderung eines - zumindest heute noch - suboptimalen Softwareproduktes auf der Anbieterseite oligopolistische oder sogar monopolistische Strukturen zu fördern.
3. Unterhalb der graphischen Programmoberfläche tut sich so einiges. Produkte wie CD Answer, Folio Views und MS Multimedia Viewer erlauben schon heute, datenbank- bzw. textübergreifend zwischen CD-ROM-Anwendungen verschiedener Anbieter zu springen, sofern diese unter derselben Software laufen und der Anwender über möglichst viel Laufwerkskapazität (CD-ROM oder Festplatte) verfügt. Für CD Answer (DOS, Windows, Unix, Apple), CD Answer Hypertext (DOS, Windows) und den Microsoft Viewer (Windows) existieren Schnittstellen, die es erlauben, mit Hilfe von Programmiersprachen oder Autorensystemen auf die Datenbestände zuzugreifen. (So weit ist Folio Views noch nicht.) Praktisch alle Softwareanbieter geloben, ihre Datenbankprodukte konsequent auf eine plattform-, d. h. betriebssystemübergreifende Client/Server-Architektur umzustellen, die eine umfassende anbieterübergreifende Connectivity erlauben wird.

Angesichts der faktischen Aussichtslosigkeit einer Einigung auf eine bestimmte Abfragesoftware jedenfalls zum jetzigen Zeitpunkt eröffnen sich für die Informationsanbieter und Verlage zwei Szenarien, die sich gegenseitig durchaus nicht ausschließen:

Szenario 1: Es gibt Softwarehäuser, die daran arbeiten, die Dualität zwischen nicht-linearem und linearem Datenbankkern mit einem umfassenden Produkt zu überwinden. Da die komplette Funktionalität einer entsprechenden Abfragesoftware verwirrend wäre, geht der Trend zur "konfigurierbaren Funktionalität". Damit ist gemeint, daß der Verlag im Produktionsprozeß entscheidet, wieviel und vor allem welche Funktionalität er für das konkret geplante elektronische Produkt benötigt. Der Anwender wird dann nur die Funktionen, die für den betreffenden Datenbestand Sinn machen, in der Abfragesoftware vorfinden.

Szenario 2: Die Informationsanbieter und Verlage gehen daran, die Schnittstellen ihrer elektronischen Produkte offenzulegen. Softwareanbieter, die hier technische oder lizenzrechtliche Barrieren aufbauen, werden von den jeweils interessierten Verlagen dazu bewegt, diese Barrieren abzubauen. Nach und

nach werden alle elektronischen Produkte für gegenseitige Zugriffe geöffnet. Nachweis- und Volltextdatenbanken, Hypertext- und Multimediasysteme sowie buchkompatible Produkte bis hin zu den hier noch gar nicht erwähnten Lexika, Lernprogrammen, Berechnungsprogrammen und Expertensystemen werden in einigen Jahren so durchlässig sein, wie es unter Windows heute schon die gängigen Büroprogramme (Textverarbeitung, Tabellenkalkulation etc.) sind.

Fazit

Wer im Begriff ist, in das textorientierte elektronische Publizieren einzusteigen oder eine bereits existierende Produktpalette zu erweitern, sollte sehr genau prüfen, welcher Textgattung seine Daten zuzurechnen sind. Handelt es sich um Kataloge, Verzeichnisse oder Fachzeitschriften? Dann empfiehlt sich eine marktgängige Volltextdatenbank. Soll ein originär elektronisches Informationsprodukt erstellt werden? In diesem Fall ist wahrscheinlich ein klassisches Hypertextsystem das richtige Tool. Geht es schließlich um die Umsetzung sequentiell strukturierter Bücher? Dann sollten die Leipziger Empfehlungen zur Umsetzung von Büchern auf elektronische Medien zu Rate gezogen werden.

Häufig wird sich zeigen, daß nicht alle Texte, die zur elektronischen Veröffentlichung anstehen, unter der gleichen Abfragesoftware realisiert werden können. Nur bei Lexika, Nachschlagewerken und Anthologien dürfte es regelmäßig unerheblich sein, ob Datenbanksoftware oder buchkompatible Software (jeweils mit Volltextsuche und Hypertextfunktion) zum Einsatz kommt.

Anschrift des Verfassers:

Dr. Rainer Dechsling
Leiter des Lektorats für elektronisches Publizieren (LeP)
in der juristischen Abteilung des Verlages C. H. Beck
Wilhelmstr. 9
80801 München
Tel.: (089) 3 81 89-423; Fax: -398

Sylvia Springer

Erfahrungen in der Produktion von elektronischen Diensten

Die Pestalozzi-Volltextdatenbank

Vorwort

Die Erstellung der vorliegenden Volltextdatenbank mit den Werken und Briefen Johann Heinrich Pestalozzis ist die "Geschichte der Mutation" einer großen, komplexen Buchausgabe in eine CD-ROM. Da bislang ein solcher Weg in dieser Vielschichtigkeit noch nicht beschritten wurde, ist der Text bewußt als detaillierter und praxisorientierter Erfahrungsbericht verfaßt, der auch nicht verschweigt, aus welchen Erfahrungen die Autoren erst im Laufe des Projektes klug wurden. So kann er möglicherweise andere, die Ähnliches planen, sachdienlich unterstützen.

Die Pestalozzi-CD-ROM ist, was das Konzept betrifft - inhaltlich, strukturell wie auch technisch -, das Ergebnis von rund 7 Jahren "Phantasie und Arbeit": **"Phantasie"** im Sinne von "Mut zum Träumen", von hochgesteckten Zielen, von Vorstellungskraft des vielleicht auch noch Möglichen; **"Arbeit"** im Sinne von Erprobung vorhandener Möglichkeiten, Evaluation von Software-Angeboten und Anpassung des Text-Grundstocks an die Bedingungen von Recherchierbarkeit.

Diese CD-ROM ist, was ihre eigentliche Realisierung anbetrifft - Aufbereitung und Strukturierung des Volltextes im Hinblick auf die ausgewählte Software, Programmierung und Herstellung -, das Ergebnis intensiven Nachdenkens und Arbeitens in einem Zeitraum von ca. anderthalb Jahren. Der Entschluß zur Umsetzung des Projektes mit der Firma Makrolog in Wiesbaden unter der Retrieval-Software CDIS fiel im Dezember 1992. Bis dahin war der Text lediglich gescannt und orthographisch vereinheitlicht. Im Juni 1994 war die CD in der DOS-Version fertiggestellt; die Versionen unter Windows und Macintosh werden Ende des Jahres erwartet.

Ohne den vorlaufenden Raum des Träumens und Erprobens wäre diese letztendlich sehr schnelle Realisierung nicht möglich gewesen. Ohne die intensive Konzentration auf die Umsetzung in den besagten 18 Monaten wäre der Traum "Pestalozzi-CD-ROM" nicht in Erfüllung gegangen. Phantasie und Arbeit bleiben für ein solches Pilot-Unternehmen untrennbar miteinander verwoben.

Ich will im folgenden den Vorlauf - die Entwicklung der Idee und die Evaluation der Möglichkeiten - darlegen; im Anschluß daran möchte ich die schrittweise Verwirklichung in der anderthalbjährigen Arbeitsphase beschreiben und die Resultate am fertigen Produkt demonstrieren. Dabei möchte ich versuchen, Überlegungen und Arbeitsschritte, so weit es geht, zu generalisieren.

Der Rahmen des Projekts

Mitte der achtziger Jahre übernahm am Erziehungswissenschaftlichen Institut der Heinrich-Heine-Universität in Düsseldorf Prof. Dr. Leonhard Friedrich - heute Professor an der Friedrich-Schiller-Universität in Jena - die Entwicklung eines dringlichen Projektes im Rahmen der Pestalozzi-Forschung: Es ging um die Erstellung von Gesamtregistern zur Kritischen Gesamtausgabe der Werke und Briefe Pestalozzis.

Diese Gesamtausgabe stellt ein in Umfang und Konzept in den Erziehungswissenschaften herausragendes Textcorpus dar. Sie wurde 1927 - im 100. Todesjahr Pestalozzis - begonnen und umfaßt heute 28 Werk- und 13 Briefbände von jeweils etwa 500 Druckseiten Umfang. In den nächsten Jahren wird sie noch um einige Ergänzungsbände mit bislang unveröffentlichten Texten und Fragmenten sowie mit Briefen an Pestalozzi erweitert werden.

Der Aufbau der einzelnen Bände ist in beiden Abteilungen - Werke und Briefe - im allgemeinen chronologisch. Jeder Band enthält vier Anhänge, nämlich einen textkritischen Apparat mit Lesarten und Varianten, Sacherklärungen zu Hintergründen von Textpassagen und/oder angeführten Personen, Erläuterungen typisch schweizerischer sowie veralteter Begriffe und ein - freilich lückenhaftes - Orts- und Personennamen-Verzeichnis. Außerdem existiert mittlerweile ein erster Band der Gesamtregister mit Verzeichnissen aller Personen- und geographischen Namen sowie diversen Unterverzeichnissen (vgl. Abb. 1: Übersicht über den Aufbau der Kritischen Pestalozzi-Ausgabe).

Die Herausgeber und Bearbeiter der einzelnen Bände hatten sich für eine "konservative Behandlung" der Texte "im Interesse strenger Wissenschaft-

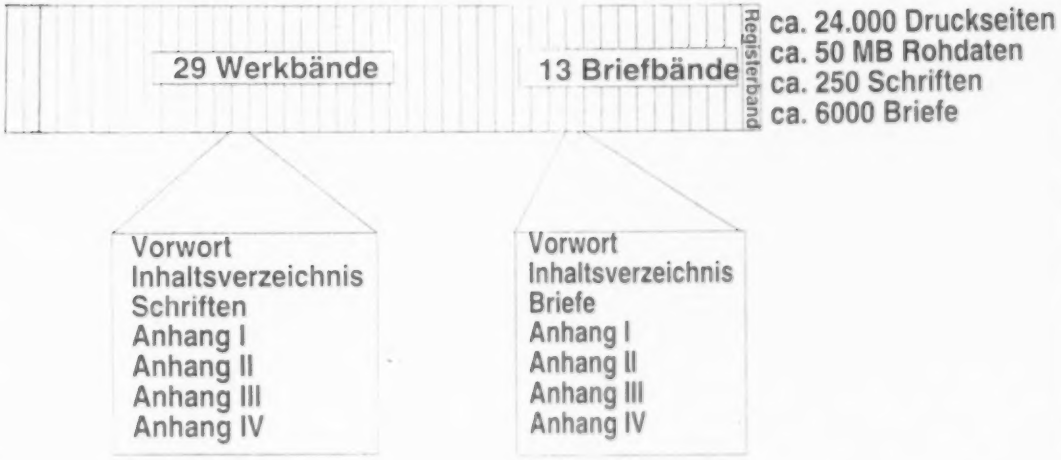


Abb. 2: Übersicht über den Aufbau der Kritischen Pestalozzi-Ausgabe

lichkeit"[1] entschieden, worunter insbesondere das Bemühen zu rechnen ist, alle orthographischen Eigenheiten Pestalozzis zu erhalten - ein Umstand, der uns bei der Erstellung der Volltextdatenbank noch stark beschäftigen sollte.

Die vorliegenden Texte beinhalten ein breites Spektrum von Fragestellungen, die über die Erziehungswissenschaft und die Lehrerbildung hinaus beispielsweise für Literaturwissenschaftler, Kulturhistoriker, Juristen und Philosophen von Interesse sind. Doch sind diese Informationen weit gestreut über ein Werk mit einem Umfang von 258 Schriften und 6252 Briefen auf rund 21000 Druckseiten; ohne umfassende und wiederum differenzierte Register ist eine gezielte, systematische und lückenlose Erschließung des Gesamtwerks - wenn diese nicht wenigen Spezialisten vorbehalten bleiben soll - ausgesprochen schwierig. Der Leser benötigt in der Regel Hilfsmittel, wie thematisch eingegrenzte Sekundärliteratur, und konzentriert sich auf ausgewählte Schriften, "Klassiker" wie den "Stanser Brief", den Volksroman "Lienhard und Gertrud" und die anthropologischen "Nachforschungen über den Gang der Natur in der Entwicklung des Menschengeschlechts".

Aus diesem Grunde übernahm Prof. Friedrich - unterstützt aus Mitteln der Deutschen Forschungsgemeinschaft - an seinem Lehrstuhl die Aufgabe, ein Gesamtsachregister sowie diverse Gesamtnamenregister, dazu etliche Unterverzeichnisse, u.a. eines der von Pestalozzi bestellten Bücher, zu erarbeiten.

Angesichts des Umfangs des Werkes, bei dessen Auswertung allein aufgrund von Lektüre - also "traditionelle" Register-Erstellung mit Hilfe von Zettelkästen - ein beträchtlicher Ausfall- und Fehlerquotient unvermeidlich bliebe, schienen die Erfassung des Textes auf EDV und die Nutzung von Retrievalsoftware den sinnvollsten Weg zu bieten. Allerdings lagen kaum Erfahrungen im geisteswissenschaftlichen Bereich vor, jedenfalls nicht mit nichtrelationalen Datenbanken.

Von der optimalen Methode der Texterfassung über die geeigneten Speichermedien bis zur adäquaten Software mußte alles durchdacht und evaluiert werden.

Die DFG schlug in einem Begleitbrief zum Bewilligungs-Bescheid die Kontaktaufnahme mit einem Literaturwissenschaftler und EDV-Experten vor, der sich insbesondere auf das neuartige Speichermedium CD (CD-WORM und CD-ROM) spezialisiert hatte. Dies könne möglicherweise die geeignete Grundlage bei der Registererstellung bieten, da es den gescannten Text in nur einer Datei präsentieren könne.

Aus Kostengründen schied dieses Medium als reine Zwischenstufe aus. Es blieb allerdings die Überlegung, längerfristig eine Publikation des Volltextes

[1] Vgl. die Vorrede der Herausgeber zur Gesamtausgabe in: Johann Heinrich Pestalozzi: Sämtliche Werke, Band 1, Berlin 1927, S.V.

auf CD-ROM im Auge zu behalten. Zunächst einmal galt es, die 42 Bände zu scannen und geeignete Hard- und Software für das Retrieval zu finden.

Die elektronische Erfassung der Texte

Sowohl der technische Stand als auch die Preise von Scanner-Hardware und OCR-Software können heute mit 1987 nicht mehr verglichen werden. Wir waren genötigt, das Einlesen der aufgeschnittenen Bände an Institutionen mit guter Ausrüstung zu delegieren - an das Universitätsrechenzentrum in Tübingen und die Gesellschaft für wissenschaftliche Datenverarbeitung in Göttingen, die über KDEM-Omnifont-Lesemaschinen der Firma Kurzweil verfügten, die für unseren Zweck - das Scannen einer sehr komplizierten Textvorlage mit verschiedenen Fonts und einer komplexen Struktur - die geeignetste zu sein schien. Von der DFG finanzierte Hilfskräfte bewerkstelligten das Scannen und Umsetzen in einer Durchschnittsrate von 13 Seiten in der Stunde (i. e. pro Seite ca. 4,7 Minuten). Ausgespart wurden reine Images wie Abbildungen und Faksimiles.

Etwas problematisch war die Umsetzung durch die amerikanische Software im Hinblick auf Umlaute und "ß", die eine Sondercodierung erhielten, die wir später zurücksetzen mußten.

Es war wichtig, bei der elektronischen Erfassung die Seiten- und Zeilenbezüge zur Druckvorlage zu erhalten, sollte die elektronische Version doch buchbezogene Register erstellen helfen. So wurden sogenannte "feste" Zeilen- und Seitenumbrüche gesetzt.

Erhalten blieben auch die Kolumnentitel mit Seitenzahl und die Zeilennummern neben dem Text (Fünfersprünge), denn die Recherche brauchte ja Anhalte, insbesondere im Falle des Überspringens von leeren Seiten.

Den verschiedenen Schrifttypen der Kritischen Ausgabe, deren Verwendung jeweils besondere Bedeutung zukommt, wurden verschiedene Fonts zugeordnet; Kolumnentitel und Zeilennummern unterscheiden sich ebenso wie Überschriften in der Schriftgröße (Klein- resp. Großdruck). Die Anhänge sind im Gegensatz zum Haupttext zum Teil in Fraktur gesetzt, Erläuterungen der Herausgeber innerhalb des textkritischen Apparates sowie Ergänzungen von Wortfragmenten typographisch ebenso abgesetzt wie von Pestalozzi gestrichene Textteile (vgl. Abb. 2: Beispiel für eine Buchseite aus dem textkritischen Apparat).

Die Bände wurden in jeweils zwei Dateien (Text- und Anhangteil) gespeichert; insgesamt umfaßt der elektronische Textbestand etwa 55 Megabyte.

Textkritik 339

S. 120 Z. 13—14 eine unschedliche, aber unangenehme pflanze
S. 120 Z. 31 gewähren lassen? fehlt
S. 120 Z. 38 werden fehlt
S. 121 Z. 3 Vernachlässigung die drei letzten Silben, mit denen die neue Zeile hatte beginnen müssen, hat J. vergessen
S. 121 Z. 9 an fehlt
S. 121 Z. 24 Solange ich ⟨mit⟩ Munterkeit
S. 121 Z. 26 hat fehlt
S. 121 Z. 31 Ordnung ⟨und entwikleter Heiterkeit⟩ Ruhe
S. 121 Z. 37 deiner Übungen nachträglich eingefügt
S. 122 Z. 2 auftretten auftreten
S. 122 Z. 3—4 Urtheile, ⟨nur⟩ *meistens* in Sachen
S. 122 Z. 11—12 Ich sage, um diese Wahrheit brauchbarer zu haben, schwecher: nachträglicher Zusatz
S. 122 Z. 22, 26 die Anführungsstriche fehlen
S. 122 Z. 36 eure im Text scheint fure zu stehen, was unverständlich ist
S. 122 Z. 37 bilden ⟨O daß ich es nie vergesse⟩ forteilen ⟨d⟩ ⟨nicht⟩ immer th⟨etig syn⟩*un*, immer entwikeln, ⟨aber nur⟩ aber
S. 122 Z. 39—S. 123 Z. 1 Zufrieden mit dem — der Wahrheit alles. Der Satz ist am Rande nachgetragen, ohne Angabe, wo innerhalb der Aufzeichnungen zum 14. Februar er einzufügen sei.
S. 123 Z. 3—4 übe, ⟨ist mir heute oft⟩ *veranlasste* folgendes überleg⟨endes Gespräch⟩ *ungen*
S. 123 Z. 6 herrührt fehlt
S. 123 Z. 31 f. vide [Herrn Johann Jacob] R[ousseaus, Bürger zu Genf, Aemil, oder Von der Erziehung. Aus dem Französischen übersetzt, und mit einigen Anmerkungen versehen. Erster Theil, Berlin, Frankfurt und Leipzig, 1762, S.] 137. Vgl. Anhang II S. 393.
S. 123 Z. 33 hinzusezt hinzusezen
S. 124 Z. 16 sich fehlt
S. 124 Z. 14—21 Urtprüngliche, durch einen aufgeklebten Zettel formgierte Fassung: Miten in deiner lichten Führung ⟨wan dem Sohn an deinem Arm⟩ ⟨*wann dir*⟩ *um weiten Hörsall der Natur*, *in Berg und Thal* ⟨irrst⟩ ⟨wo du ihn lehrest auf der⟩ ⟨lasse jede sich darbietende⟩ Gegenstände Lehre syn. Aber wann du ihn jez auch etwas [anderes] lehrest als die gegenwertige Natur dir darbeut, so laß die Freude uber den Fortgang
S. 124 Z. 22 reißen, ⟨ihme selbst in den⟩ wan ⟨frappirende⟩ Gegenstände
S. 124 Z. 24—27 lassen. ⟨Wan miten in der Sprachübung das Roß wiehert, der Vogel singt,⟩ ⟨der⟩ ein ⟨wan es⟩ ⟨Land und Berg ist sein Hörsall, Lehrstube jeder Auftrit in diesen, und in diesem weiten Hörsall⟩ Soll ⟨du⟩ ⟨wann es vill ist, wann er⟩ er ⟨wussen⟩ *es empfinden*, ganz empfinden, daß die Natur ⟨zwei Worte reden⟩ *hier lehren* soll und du nur diese und still *mit* die folgende Kunst ⟨auftrinnen darfst⟩ *vast neben hinschleichen sollst*, wann der Vogel
S. 125 Z. 1—2 so werden so sie
S. 125 Z. 3 gemacht fehlt

Abb. 2: Beispiel für eine Buchseite aus dem textkritischen Apparat

Herstellung von Recherchierbarkeit der Texte

Die Pestalozzi-Gesamtausgabe wurde, wie bereits erwähnt, nach strengen textkritischen Maßgaben herausgegeben. Für die Recherchierbarkeit eines Textes stellt es wiederum ein bedeutendes Hindernis dar, wenn Schreibvarianten verwendet werden, die so ungewöhnlich sind, daß man darauf kaum kommen könnte. Genau dies ist bei Pestalozzi der Fall. In der Kritischen Ausgabe findet man nicht nur solche für das 18. und beginnende 19. Jahrhundert typischen Schreibvarianten wie "th" statt "t" oder "ey" statt "ei". Solche Möglichkeiten ließen sich beim Recherchieren leicht einbeziehen. Man stößt aber auch auf Abweichungen wie "Teütschland" für "Deutschland" und "Jde" für "Idee". Daß hiermit nicht gewollte orthographische Finesse dokumentiert wird, belegt Pestalozzi selbst, der von sich bekennt, er habe seine Texte "ungewaschen und ungekämmt ins Publikum geworfen"[2]. Insofern scheint orthographische Vereinheitlichung kein Sakrileg zu sein; immerhin findet man sie auch innerhalb der Kritischen Ausgabe bei zu Lebzeiten veröffentlichten Texten. Das heißt, daß auch hier der korrigierende Eingriff des Verlags erwünscht war.

Die gescannten Texte mußten ohnehin auf Einlesefehler hin kontrolliert werden. Es gab darunter ganz typische wie "umd" für "und" und "jhr" für "ihr", durch die Frakturschrift bedingte wie "f" für "s", aber auch eine Vielzahl von nicht voraussagbaren, die Folge der teilweise nicht sehr guten Vorlage waren. Die bereits erwähnte Variantenvielfalt hätte es zum Wiederherstellen des Zustandes der Textvorlage erforderlich gemacht, den gescannten Text Wort für Wort mit der Druckvorlage zu vergleichen. Auf diesem Wege hätte dann zugleich jeder ungewöhnlichen Schreibweise die heute gültige verdeckt hinterlegt werden können, um das Retrieval aller Varianten zu gewährleisten. Dieser Abgleich von 21000 Druckseiten war ökonomisch nicht realisierbar und scheint aus den bereits diskutierten Gründen auch nicht zwingend.

So entschied sich die Projektleitung für die orthographische Vereinheitlichung des Textes nach heutigem Standard, sozusagen also für die "kosmetische Aufbereitung der ungepflegten Vorlage". Zur Dokumentation des Grades der Abweichung ist der erste Band der Werkausgabe als Faksimile der CD-ROM beigegeben. Von jeder Textseite aus kann das entsprechende Bild aufgerufen werden. Aus Speicherplatzgründen wurde dieses Verfahren noch nicht für alle Bände durchgeführt, ist allerdings für die nächste Auflage vorgesehen.

Die Korrekturarbeiten ließen sich mit herkömmlichen Rechtschreibprogrammen nicht bewerkstelligen. Dem äußerst vielseitigen, kreativen Wortschatz Pestalozzis waren jene Programme nicht gewachsen. Ihre Erweiterung wäre

[2] Vgl. Johann Heinrich Pestalozzi: Sämtliche Briefe. Zürich 1971, Bd. 13, S.300.

extrem aufwendig gewesen und hätte an Umfang vermutlich die damaligen Arbeitsspeicher überfordert. Darüber hinaus wären semantische Fehler, etwa durch falsches Einlesen erfolgte Buchstabenverwechslungen, überhaupt nicht gefunden worden.

Also wurde der komplette Text am Bildschirm korrigiert, bandweise, allerdings unter Zuhilfenahme von Makros mit "Suche und Ersetze"-Befehlen für sehr häufige Schreibvarianten wie "gieng" ("ging") und "jez" ("jetzt"). Da auf diese Weise einige neue Fehler, insbesondere bei Eigennamen entstanden, mußte beim anschließenden Lesen gebührend darauf geachtet werden. So wurde z. B. die schweizerische Stadt "Thun" ein "Opfer" der "th-Bereinigung".

Ausgeklammert werden mußten auch die englisch- und französischsprachigen Texte; aus jedem englischen "that" wäre sonst ein simples "tat" geworden. Selbstverständlich erfaßten unsere "Korrekturen" nur echte orthographische Varianten. Semantisch bedeutsame Abweichungen wie "Ahnung" und "Ahndung", "Fortschritt" und "Vorschritt" blieben erhalten.

Beim Korrekturlesen wurden Leistungen von durchschnittlich ca. 6 Seiten pro Stunde erreicht. Da mehrere Personen gleichzeitig daran arbeiteten, waren regelmäßige Besprechungen zur Gewährleistung gleichsinniger Richtlinien erforderlich. Nicht immer gab der Duden einhellige Auskunft.

Zum Abschluß dieses aufwendigen, aber notwendigen Arbeitsganges wurden mit Hilfe des Programms "Trenn-Ex" die Silbentrennungen an Zeilenumbrüchen aufgehoben, um auch die betreffenden Worte als Einheiten recherchierbar zu machen.

Hardware und Retrieval-Software

Große Festplatten auf PC zur Bearbeitung des Textes von 55 Megabyte in einem Stück waren zum damaligen Zeitpunkt nur zu Preisen zu bekommen, die im Rahmen der Projektmittel unerschwinglich waren. Da dieses Thema heute nicht mehr relevant ist, will ich es nicht diskutieren, sondern nur vermerken, daß wir als für unsere Zwecke annähernd optimale Lösung PC mit Wechselspeicherplatten von 30 Megabyte wählten. Auf einer Platte war Platz für ein Drittel des Textes in indexierter Form, so daß wir in drei Arbeitsgängen unsere Recherchen für die Register durchführen konnten.

Als Indexierungs- und Retrieval-Software entschieden wir uns für das Programm WordCruncher, in den USA zur Bearbeitung umfangreicher Mormonen-Genealogien entwickelt und auf dem deutschen Markt für eine Textmenge unserer Größenordnung gut geeignet. Viele andere Programme hätten

eine Zerlegung des Textes in sachlich unsinnige Einheiten erforderlich gemacht.

Von der Register-Erstellung zur CD-ROM-Konzipierung

Die Pestalozzi-Gesamtausgabe lag nun elektronisch erfaßt und orthographisch vereinheitlicht, also leicht recherchierbar vor. Das Einfachste wäre es gewesen, die CD-ROM mit dem Programm WordCruncher auf den Markt zu bringen. Doch so sehr dieses den Text als ganzen erschließen half - mit hoher Geschwindigkeit Fundstellen als Keyword-in-Context präsentierte -, auch das Zusammenstellen und Abspeichern von Suchlisten ermöglichte, so unerfreulich langsam wurde es bereits bei der Vorne-Trunkierung. Doch vor allen Dingen schien es aufgrund strenger Restriktionen des amerikanischen Herstellers den deutschen Vertreibern nicht möglich zu sein, einige Veränderungen im Programm vorzunehmen, wie sie uns für eine CD-ROM-Ausgabe wichtig erschienen, wenn sie dem Buch in einigen Punkten überlegen sein sollte; insbesondere ist darunter die Verknüpfung von Hypertexten zu fassen.

Die Arbeit mit WordCruncher lehrte uns durchaus, welche erstaunlichen Retrieval-Möglichkeiten bereits bestanden. Gerade auch quantitative Erhebungen (Worthäufigkeiten insgesamt und in einzelnen Texten) eröffneten interessante Forschungsperspektiven. Zugleich weckten diese Retrieval-Erfahrungen aber auch Vorstellungen von einem noch leistungsfähigeren Programm.

Dieses sollte z. B. verschiedenste Hypertext-Anbindungen ermöglichen, Zusatzinformationen bieten, wie die als Images gescannte Originalversion und diverse Abbildungen, sowie systematische Zugriffe ermöglichen.

Wir waren gefragt, unsere Vorstellungen in einem Pflichtenheft klar zu umreißen; wir waren außerdem gefragt, infrage kommende Software-Anbieter zu ermitteln, ihre Produkte zu prüfen und in der Konfrontation mit unserem Pflichtenheft die Möglichkeit der Zusammenarbeit zu erwägen.

Als Mischtyp aus großen Mengen von Fließtext einerseits, zahlreichen Hypertext-Verknüpfungen andererseits, aus Texten und aus Bildern, mit Bedarf an Volltextrecherche ebenso wie an feldorientierter Suche, konnte für die Pestalozzi-Volltextdatenbank das Angebot vorliegender Autorensysteme die Anforderungen an Software nicht hinreichend erfüllen. Die Unterstützung durch versierte Programmierer war erforderlich, wenn die Realisierung dem Werk Pestalozzis und den bestehenden Forschungsinteressen angemessen sein sollte.

Das Pestalozzianum in Zürich unter der Leitung von Prof. Dr. Hans Gehrig erklärte sich zur Finanzierung des Projekts bereit; bei dieser Institution liegen die Vertriebsrechte der CD-ROM.

Pflichtenheft und Software-Evaluation

So galt es, eventuellen Software-Partnern zum einen eine möglichst präzise Vorstellung vom vorliegenden Material zu vermitteln, zum anderen die gewünschten Zugriffsweisen zu beschreiben, Verknüpfungswünsche zu benennen, Ablagemöglichkeiten anzugeben und die technischen Anforderungen zu klären. Es entstand ein neunseitiges Pflichtenheft, das genau diese Angaben enthielt. Vom heutigen Standpunkt aus wirkt es zumindest unvollständig. Zur damaligen Zeit stellte es einen erforderlichen Schritt auf dem Weg zur Umsetzung dar, diente es doch der Klärung von Zielen, die man erreichen wollte. Es wird hier in seinen Grundzügen kurz skizziert, um bereits die Aufgaben zu umreißen, die im Hinblick auf Textaufbereitung bevorstehen sollten.

Als *erste Aufgabe* wurde der vorliegende Datenbestand beschrieben: die eigentlichen Textteile, der textkritische und der sacherklärende Anhang, die kumulierten Register und Worterklärungen, außerdem der in Faksimile-Images vorliegende erste Band und Abbildungen von Zeitgenossen, Lebensstationen und Wirkungsstätten Pestalozzis. Unter die gewünschten Zugriffsweisen wurde zunächst "Lektüre" gerechnet. Buchidentische Verwendungsmöglichkeiten wie das "Aufschlagen" beliebiger Seiten und fortlaufendes "Blättern" sollten gewährleistet sein. Des weiteren war natürlich Volltextrecherche gewünscht, unterstützt durch die Auswahlmöglichkeit aus einem Gesamtwortindex.

Selbstverständlich sollten die Boolschen Verknüpfungen "und", "oder" und "nicht" bei der Volltextrecherche verwendet werden können, von bis zu drei Worten in einem frei wählbaren Umfeld, z. B. der Einheit "Satz".

Fundstellen sollten als Textausschnitt (Keyword-In-Context/KWIC) oder als Kurztitel mit buchidentischer Fundstellenanzeige, überdies auf Wunsch in chronologischer Sortierung, angezeigt werden. Es war auch daran gedacht, Informationen systematischer Art verdeckt im Text zu hinterlegen und recherchierbar zu machen. Hypertext-Anbindungen sollten zwischen den Textteilen und den beiden zugehörigen Anhängen, zwischen Texten sowie Anhängen und den erläuternden Registern, zwischen den Texten und der als Bilder gescannten Originalversion sowie anderen Abbildungen vorgenommen werden. Zum Export sollten sowohl beliebig markierte Textpassagen als auch Fundstellenlisten bereitstehen.

Unter den technischen Wünschen spielte die Systemanforderung Macintosh eine erhebliche Rolle, da das Schweizer Bildungswesen überwiegend unter diesem System arbeitet, dem die Pestalozzi-CD-ROM selbstverständlich auch zur Verfügung stehen sollte.

Die *zweite Aufgabe* bestand in der Evaluation der angebotenen Programme. Sie wurde etwa zeitgleich betrieben. Durch Besuche einschlägiger Messen

BEST COPY AVAILABLE

und die Auswertung von Fachzeitschriften konnte ein Überblick über potentielle Kooperationspartner gewonnen werden. Relevant war u.a. die bisher vorgelegte Produktpalette, wenn auch eine Volltextdatenbank wie die geplante in dieser Komplexität noch nicht existierte. Fünf Firmen, die ernsthaftes Interesse an der Zusammenarbeit signalisiert hatten, wurden gebeten, im Hinblick auf das mittlerweile fertiggestellte Pflichtenheft eine Vorführung ihres Programmes zu bieten, wenn möglich unter Einbeziehung eines kleineren Pestalozzi-Textes, um zumindest im Ansatz die Realisierbarkeit unserer Vorstellungen zu dokumentieren.

Für die Auswertung dieser Vorführungen wurde eine Checkliste erstellt, die sich streng an den Forderungskanon des Pflichtenheftes hielt, allerdings auch bereits eine Kalkulation der Kosten vorsah. Der Evaluierungsprozeß erfolgte in seiner Entscheidungsphase in enger Zusammenarbeit mit dem Pestalozzianum.

Die Probe-Indexierungen des Pestalozzi-Textes waren generell nicht sehr überzeugend. Bereits bei kleinen Textmengen zeigten sich strukturelle Schwierigkeiten in der Textanordnung und -zuordnung in der Buchausgabe, die auf einen beträchtlichen manuellen Aufwand in der Textaufbereitung hinzudeuten schienen.

Die Firma MAKROLOG mit der in Israel entwickelten Indexierungs- und Retrieval-Software CDIS erhielt den Auftrag.

Erste Schritte der Zusammenarbeit

Die Demo-Version unter CDIS im Dezember 1992 hatte zwar wiederum deutlich gemacht, daß wir nicht auf Erfüllung aller unserer Vorstellungen würden bestehen können, daß aber andererseits komplexere Verknüpfungen und Strukturierungen möglich sein könnten, als sie im Pflichtenheft benannt worden waren.

MAKROLOG erstellte darum im Anschluß an die Diskussion der Demo-Version ein angepaßtes und ergänztes Pflichtenheft. Auf dieser Grundlage wurde eine Festlegung der von allen beteiligten Seiten zu leistenden Arbeiten vorgenommen: einerseits die von uns noch am Text vorzunehmenden Arbeiten, andererseits die von der Firma programmierbaren Aufbereitungsschritte, schließlich auch der ständige Dialog, um "Fehlpässe" zu vermeiden und Hand in Hand zu arbeiten.

Die Aufgaben, die für die Projektmitarbeiter in Jena unter Prof. Friedrich mittlerweile gewechselt hatten, und in Düsseldorf unter meiner Leitung anstanden, wurden möglichst präzise beschrieben. Sie umfaßten Aufgaben, die dazu dienen sollten,

- den Text an sich vollständig recherchierbar zu machen,
- den Text in verschiedenen Einheiten wie Satz und Absatz recherchierbar zu machen,
- verständliche und präzise Fundstellenangaben bieten zu können,
- korrekte Hypertext-Anbindungen zu gewährleisten,
- systematische Zugriffe zu ermöglichen und
- ergänzende Hinweise anzubieten.

Diese Aufgaben wurden im Rahmen der Möglichkeiten im Zeitraum Anfang 1993 bis Juni 1994 erfüllt. Der Weg ihrer Erfüllung zum einen, die Ergebnisse auf der vorliegenden CD-ROM zum anderen sollen nun im zweiten Teil dieses Beitrages dargestellt werden.

Den Text an sich recherchierbar machen

Wie bereits beschrieben, war der gescannte Text orthographisch vereinheitlicht worden. Dennoch standen etliche typographische und Satz-Probleme einer vollständigen Recherchierbarkeit im Weg. Dazu zählte zunächst die Vielzahl setzerischer Besonderheiten in den Anhängen. Unvollständige Worte Pestalozzis waren durch Bearbeiter mutmaßend ergänzt; um diese Ergänzung kenntlich zu machen, stand sie in Klammern. Das wiederum unterbrach die Buchstabenfolge des Wortes und ließ es bei Rechercheanfragen nicht auffindbar sein. Bearbeiter mußten nun das Fragment wiederherstellen und dahinter das vollständige Wort in Klammern eintragen, um kenntlich zu machen, daß es sich hierbei um ein teilweise oder vollständig ergänztes Wort handelt.

Von Pestalozzi gestrichene Text-, aber auch Wortteile waren durch einen anderen Klammer-Typus gekennzeichnet. Hier gab es den parallelen Fall, daß Wörter aufgrund teilweiser Buchstabenumklammerung nicht recherchierbar waren; auch hier mußte auf das Fragment oder den Teil des Wortes, den Pestalozzi stehen lassen hatte, zusätzlich in Klammern das von ihm zuvor gestrichene Wort folgen (vgl. zu diesen Schwierigkeiten Abbildung 2).

Ein weiteres Problem stellten jetzt die Kolumnentitel und die Zeilennummern dar, die in Fünfersprüngen am Textrand standen. Die Kolumnentitel mit der ständigen Wiederholung einer Überschrift wie z. B. "Wie Gertrud ihre Kinder lehrt" erhöhten auf irrelevante Art und Weise die Gesamtzahl von Fundstellen etwa für die Verbform "lehrt"; die Ziffern führten zu unschönen Zeileneinrückungen. Da beide Angaben durch das Retrievalprogramm überflüssig wurden, das in einer Zeile außerhalb des Textfensters Kurztitel und aktuelle Zeilennummer des Cursors angibt, mußten die entsprechenden Stellen gelöscht werden. Dank der Kennzeichnung von Kolumnentiteln und Zeilenzif-

fern durch Spitzklammern als Kleindruck-Font war diese Bereinigung fast automatisch zu bewerkstelligen.

Nach den Vorarbeiten schien der Gesamtwortindex relativ brauchbar: Der Text konnte jetzt nach jedem Begriff durchsucht werden; durch verschiedene Weisen der Trunkierung war ein Wort mitsamt Flexionen recherchierbar, zumal Umlaute mitberücksichtigt werden (vgl. Abb. 3: Ergebnisbox für die Vorne- und Hinten-Trunkierung um "*rechtig*".

Die ursprünglich vorgesehene halbautomatische Lemmatisierung des Wortbestandes, um das Recherchieren auf Grundformen reduzieren zu können, war aufgrund des schlechten Resultates verworfen worden. Natürlich ist mit Hilfe von Trunkierung auch die Erhebung sämtlicher Komposita möglich, die zunächst in einer Auswahlbox angeboten werden und aussortiert werden können.

Eine Hilfestellung innerhalb des Index wurde allerdings im Hinblick auf die fremdsprachigen Texte erbracht. Der Pestalozzi-Leser hat in der Regel die deutschen Termini vor sich. Oft wird vergessen, daß beispielsweise die nur in englischer Sprache vorliegenden "Letters on Early Education" wichtige Überlegungen Pestalozzis zur frühkindlichen Erziehung beinhalten und daß z. B. in den zahlreichen Briefen an die Eltern von Schülern und Schülerinnen in der französischsprachigen Schweiz interessante Aspekte zur Beurteilung von Pestalozzis Schülerschaft verborgen sind. Aus diesem Grunde griffen wir aus den gesondert erhobenen englischen und französischen Gesamtwortbeständen alle für Pädagogik, Politik etc. sinntragenden Nomina und Verben heraus und ordneten sie den deutschen Äquivalenten zu. So findet der Nutzer, wenn er im Index z. B. "Schule" aufruft, zugleich das Angebot, mit dem Begriff "Schule" auch "école", "institut" und "school" in die Suchbox zu übernehmen. Er ist nicht genötigt dazu, findet hier aber eine Unterstützung.

Den Text in bestimmten Einheiten recherchierbar machen

Nach einzelnen Worten und Wortfragmenten konnte nun im Text als ganzem gesucht werden. Ebenso war die Wortabstandsuche innerhalb eines frei wählbaren Abstandes von einzelnen Wörtern möglich. Um beispielsweise bestimmte Zitate aufzufinden, ist präzise Wortabstandsuche ebenso sinnvoll wie etwas zur Überprüfung des Vorhandenseins von angeblich neuen Text-Funden im bereits gedruckten Bestand.

Um jedoch nach bestimmten Sachverhalten, z. B. der Rolle des "Lehrers" in der "Schule", zu recherchieren, führt in der Regel die Suchanfrage nach beiden Begriffen innerhalb sinnstiftender Einheiten wie Satz und Absatz weiter. Dies stellte uns vor die Schwierigkeit der Satzende- und Absatzende-Erkennung. Letztere wurde automatisch durchgeführt, indem überall da, wo die

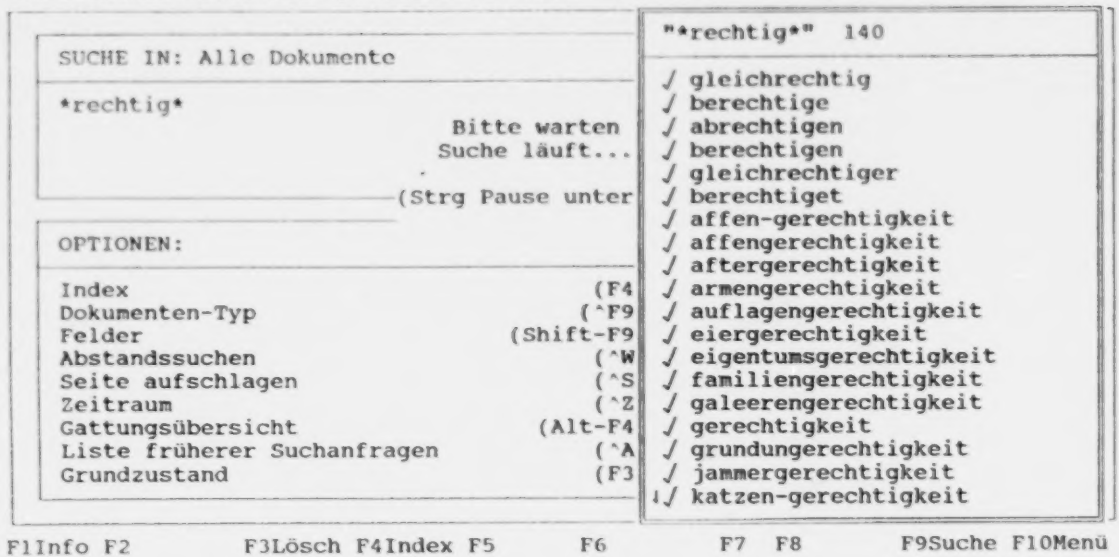


*Abb. 3. Ergebnisbox für die Vorne- und Hinten-Trunkierung um „*rechtig*"*

durchschnittliche Zeichenmenge pro Zeile um einen bestimmten Anteil unterschritten wird, eine Markierung gesetzt wurde. Der Fehlerquotient ist dabei so niedrig, daß auf eine Kontrolle des gesamten Textes verzichtet werden konnte. Allerdings hätte diese nachträgliche Arbeit erspart bleiben können durch entsprechende Markierungen der Absatzenden bereits beim Scannen.

Satzenden lassen sich deutlich schwieriger verläßlich ausmachen, da die häufigste Satzendemarkierung - der Punkt - in Texten auch andere Funktionen erfüllt, insbesondere die Kennzeichnung von Abkürzungen. Zu deren Erhebung wurde ein kleines Programm geschrieben, mit dessen Hilfe Worte vor Punkten herausgeschrieben wurden, die aufgrund bestimmter Merkmale "in Verdacht" standen, Abkürzungen zu sein, beispielsweise nachfolgende Kleinschreibung, ungewöhnliche Reihungen von Konsonanten oder die deutliche Nähe eines weiteren Punktes. Auf diese Weise wurde eine Liste mit mehreren hundert Wörtern erzeugt, die daraufhin überprüft wurde, ob es sich um Abkürzungen handelte; bei Pestalozzi fanden sich sehr eigenwillige Kürzel wie "Msgh" für "Meine sehr geehrten Herren" oder "nvbre" für "novembre". Sofern dies zutraf, wurden die Abkürzungen in der Liste aufgelöst, damit die CD-ROM zugleich ein Abkürzungsverzeichnis aufnehmen konnte. Die übrigen Begriffe wurden aus den Listen gelöscht. Nun konnte ein weiteres Programm sozusagen im Ausschlußverfahren diejenigen Punkte als Satzende-Marken dispensieren, die auf eines der Kürzel folgten. Der Fehlerquotient ist durch dieses Verfahren auch im Satzerkennungsbereich sehr niedrig.

Zwei Besonderheiten des Buchsatzes, denen die elektronische Fassung nicht ohne weiteres entsprechen konnte, waren beim Scannen ebenfalls nicht deutlich genug bedacht worden: Texte in Spalten und Texte in Fußnoten, die über mehrere Seiten fortlaufen.

Spaltentexte waren beim Scannen in der Regel nebeneinander stehengeblieben. Das warf zwei Schwierigkeiten auf: Zum einen sind gerade bei zweispaltigen Texten die Zeilen so lang, daß sie nicht in eine Bildschirmzeile passen, sondern umbrochen werden. Dann steht allerdings ein Teil des zu Spalte 2 gehörigen Textes optisch angebunden an Spalte 1; dies führt zu deutlichen Irritationen beim Lesen. Zum zweiten macht die Schreibung in Spalten am Bildschirm die Recherche in der Einheit Satz unmöglich, da das Programm die kompletten Bildschirmzeilen nach Punkten durchsucht.

Hier mußten Umsetzungen vorgenommen werden. In künstlich eingefügten "Nullzeilen", also Zeilen, deren Ziffer mit der vorangehenden Textzeile identisch gesetzt wurde, um die authentische Zeilenfortzählung nicht zu beeinträchtigen, wurde der Hinweis "Spalte 1" bzw. "Spalte 2" in den Text eingefügt. Danach wurden die Spalten hintereinander gesetzt. Die hinweisende Überschrift und die relativ schmale Zeilenbreite deuten für den Leser auf die Notwendigkeit hin, den Text der zweiten Spalte vergleichend heranzuziehen

```
                                F1 Hilfe, F10 Menu

 Tagebuch Pestalozzis über die Erziehung seines Sohnes

    19. März  Die Ausweichung des Lehrertons, Pedanterie, geniert mich.
    Wo soll ich die Grenzen zwischen Freiheit und Gehorsam,
    dessen frühe Angewöhnung im sozietätischen Leben notwendig
    ist, finden?
    [Spalte 1]
    Gründe für Freiheit.
    Jede Hemmung der Freiheit
    legt den Widerwillen
    des Hemmenden in das Herz
    der Kinder.
    Die Erfahrung zeiget,
    daß die am meisten
    gehemmten Kinder durch
    Ausgelassenheit sich die
    Hemmung ihres Willens bezahlt
    machen.
    Kinder in ihrem Willen
    zu hemmen ist ohne Reizung

PSW 1:126/14 (1774)                                   Gattung: 12
F1Info F2        F3        F4        F5Notiz F6Verweis F7↓ F8Zurück F9Suche F10Menü
```

Abb. 4: Beispiel für einen in Spalten gesetzten Text, der als solcher deutlich erkennbar ist

(vgl. Abb. 4: Beispiel für einen in Spalten gesetzten Text, der als solcher deutlich erkennbar ist).

Wo Pestalozzi nur sehr kurze Textteile, z. B. Begriffe, kontrastieren wollte, wurde eine Spaltenschreibung beizubehalten versucht; hier entfällt ja auch das Argument des auseinandergerissenen Satzes.

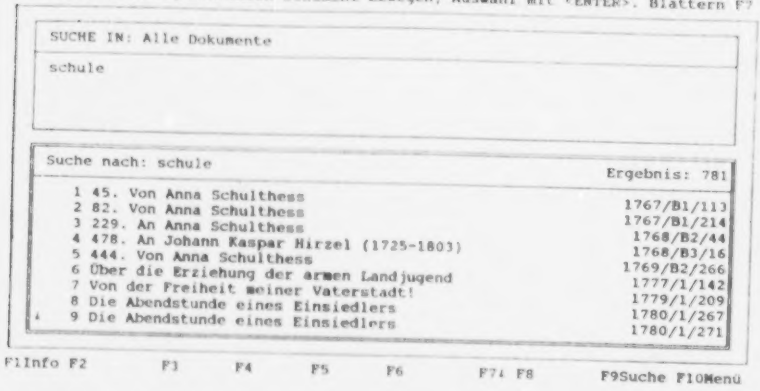
Komplex wurde das Problem bei Spaltentexten, die über mehrere Seiten hinweg laufen. Hier wurde ebenfalls jeweils eine Spalte erst vollständig aneinandergesetzt mit jeweiligen Seitenumbruchsmarkierungen, um dann die zweite hintenanzustellen, wiederum mit der betreffenden vorherigen Seite beginnend; d. h. die Seitenzahl in der Fundstellenanzeige springt automatisch zurück.

Ganz ähnlich stellt sich das Problem dar bei Fußnoten, die unten auf einer Seite beginnen und am Fuße der nächsten und möglicherweise auf weiteren Seiten fortgesetzt werden. Das heißt, der Text gehört in einen Zusammenhang, der im Buch optisch erkennbar, jedoch durch das Scannen der Seiten unterbrochen ist. Beim Scannen wurde nach dem eigentlichen Text selbst der im Buch zur optischen Absetzung angebrachte Strich nicht erfaßt, da die Zeilenzählung diesen nicht berücksichtigt, sondern fortlaufend in die Fußnote weiterzählt. Dennoch liest das Auge auf dem Bildschirm in der nächsten Zeile weiter, wo die umfangreiche Fußnote beginnt. Dieser unvorteilhafte Sprung - verbunden mit der Notwendigkeit, die Fortsetzung des Textes durch Scrollen zu suchen - wurde aufgehoben durch die Hypertext-Anbindung von Fußnoten an Textpassagen. Der Fußnotentext ist nun verborgen und stört den Lesefluß nicht.

Verständliche und präzise Fundstellenangaben bieten

Zwei Anzeigemodi wollten wir anbieten: zum einen das Schlagwort im Kontext (KWIC), mit verschiedenen Optionen der Zeilenanzahl, um jede Fundstelle eines gesuchten Begriffes direkt einsehen zu können; auf diese Weise läßt sich oft schon auf die Relevanz einer Stelle schließen; zum anderen eine Auflistung der Dokumente, in denen der Begriff vorkommt (vgl. Abb. 5: Beispiel für Fundstellenangaben als Dokumentenliste - chronologisch sortiert - und als KWIC).

Was aber ist ein Dokument? Definiert man es als Texteinheit, im Hinblick auf die es sinnvoll erscheint, das Vorhandensein eines Begriffs oder Sachverhaltes zu überprüfen, so wirft eine so vielfältige Texte enthaltende Ausgabe wie die hier vorliegende wichtige Entscheidungsfragen auf: Soll man jeden Text als Dokument auffassen oder die Texte splitten? Wenn ja, nach welchen Kriterien? Zahlreiche kurze Texte wie z. B. sämtliche Briefe, etliche Flugschriften, Aufrufe und kurze Zeitschriftenartikel sind sinnvoll als je ein Dokument



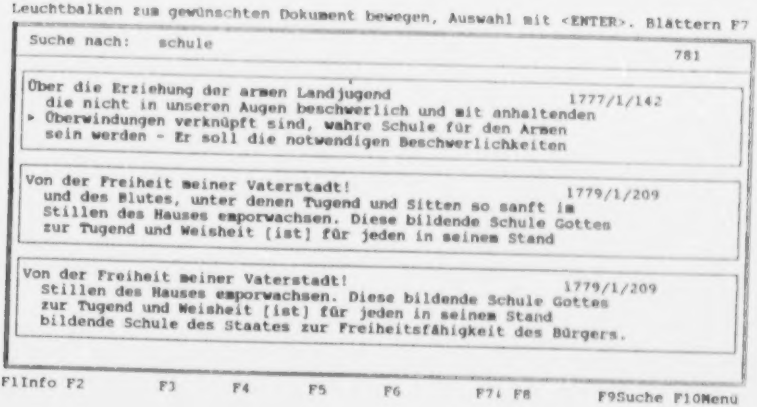


Abb. 5: Beispiel für Fundstellenangaben als Dokumentenliste (chronologisch sortiert) und als KWIC

BEST COPY AVAILABLE

aufzufassen. Die Information, ein bestimmtes Thema sei im Brief an X angesprochen, ist aufschlußreich genug; sie ist es nicht, wenn dasselbe Thema in einer über hundert Seiten langen Schrift erwähnt wird. Erfahrungsgemäß werden in einem so langen Text viele Themen angesprochen. Interessant für den Leser wird es hier, ob der gesuchte Begriff oder Sachverhalt mehrmals erwähnt wird, ob es besonders "dichte" Stellen gibt. Bietet der Text von sich aus Einteilungen an, z. B. Kapitel, wie der Roman "Lienhard und Gertrud" oder Datierungen, wie die Tagebücher, so sind Dokumenten-Splits sinnvoll dort anzusetzen. Handelt es sich jedoch nicht um strukturierten Text, so wird eine Einteilung in Seiten vorgenommen, allerdings so, daß die Einheiten Satz und Absatz nach wie vor für die Recherche zugänglich sind.

Der Sinn der Auflistung von Dokumenten besteht darin, einen Überblick über die relevanten Texte bzw. Textteile zu gewinnen. Das Splitten der Datenbank in Dokumente konnte in den Briefbänden automatisch erfolgen, da jeder Brief einheitlich mit einer Ziffer, gefolgt von einem Punkt, überschrieben ist. In den Werkbänden hingegen mußte jeder Textanfang kodiert sowie jede Splitstelle gekennzeichnet werden; lediglich für den Fall des Seitensplitting genügte ein entsprechender Code im Text als Hinweis für die Programmierer auf fortlaufendes Seitensplitting bis zum Ende-Code.

Der Kennzeichnung des Text- bzw. Dokument-Beginns kam jedoch noch eine andere wichtige Bedeutung zu: Von dieser Stelle an galt jeweils ein neuer Dokumententitel - die Anzeige, die auch in der Fundstellenliste erscheint und die bei Einsichtnahme in den Text ebenso angezeigt wird wie beim Export von Text.

Der Dokumententitel enthält folgende Angaben: den Titel des Textes bzw. eine sinnvolle Kurzform, das Datum des Erscheinens bzw. bei nicht veröffentlichten Texten der Abfassung und die Beschreibung der entsprechenden Fundstelle im Buch (Band und Seite). Letzteres ließ sich aufgrund der vorhandenen festen Seitenumbrüche und der bandweisen Archivierung relativ fehlerfrei automatisch erstellen. Die beiden erstgenannten Angaben hingegen mußten, wiederum kodiert, eingegeben werden.

Nur wenige Titel unterschritten die höchstmögliche Zeichenzahl. Nachdem alle Titel im Datenbestand gekennzeichnet worden waren, wurden sie in eine Liste geschrieben und probeweise an der kritischen Stelle geschnitten. Diese künstlichen Schnittstellen wurden durch sinnvolle Unterbrechungen ersetzt bzw. der gesamte Titel so umformuliert, daß er zwar kurz genug, aber zumindest etwas informativer hinsichtlich des Inhalts war als der reine Beginn der Überschrift. "Meine Nachforschungen über den Gang der Natur in der Entwicklung des Menschengeschlechts" war beispielsweise eindeutig zu kürzen, zumal der Text ohnehin oft nur als "Nachforschungen" bezeichnet wird. Hingegen mußte eine komplexe Formulierung wie "Skizze eines Memoire über

die Verbindung der Berufsbildung mit den Volksschulen" sinnvoll gekürzt werden zu "Verbindung der Berufsbildung mit den Volksschulen".

Für die Briefe wurden Adressatenzeilen geschaffen. Die vorliegenden Angaben - jeweils im Anschluß an die Briefnummer - waren sehr uneinheitlich, orientiert an der jeweils von Pestalozzi verwandten Anrede. Beispielsweise für seinen Freund Georg Heinrich Ludwig Nicolovius im preußischen Innenministerium finden sich Adressatenzeilen wie "An Nicolovius", "An Seiner Hochwohlgeboren", "An Georg Heinrich Ludwig Nicolovius" oder "An Monsieur". Um jedoch in der Fundstellenliste eine informative Anzeige bieten zu können, wurden diese Anreden sämtlich vereinheitlicht und zwar auf Vor- und Nachname sowie - bei gleichlautenden Namen - die Lebensdaten - selbstverständlich nicht im authentischen Text, sondern - wie bei den Kurztiteln für die Textbände - in einer eigens erzeugten Liste. Auch hier war es notwendig, im Falle besonderer Längen sinnvoll abzuschneiden, etwa bei der Nennung von mehreren Adressaten oder einer sehr ausführlich betitelten Behörde.

Das Jahr der Abfassung bzw. des Erscheinens war in der Regel aus den Anhängen erschließbar; es wurde allerdings über die hier angezeigte Jahreszahl hinaus so weit wie möglich eine Präzisierung bis zum Monat hin vorgenommen. Diese erfüllt eine wichtige Funktion für systematisch eingeschränkte Recherchen, worauf an späterer Stelle hingewiesen wird (vgl. Abb. 6: Beispiel für zwei sinnvolle Kurztitelangaben aus einem Werk- und einem Briefband).

Doch auch im Rahmen der Möglichkeiten beim Anzeigen von Fundstellen kommt ihr Bedeutung zu, denn der Benutzer kann sich die Fundstellen chronologisch sortieren lassen. Um z. B. die Genese einer Idee zu verfolgen, ist eine solche Anzeigevariante durchaus von Vorteil. Wenngleich Werk- und Briefbände im Prinzip chronologisch aufgebaut sind, gibt es Abweichungen von Bedeutung, etwa die drei Fassungen von "Lienhard und Gertrud", die in den Bänden 2-6 stehen, obwohl Jahre des Schaffens dazwischenliegen, oder Entwürfe und Fragmente, die deutlich früher entstanden als der letztendliche Text, ihm aber in der Ausgabe unter- und damit auch nachgeordnet werden. Außerdem lassen sich beim chronologischen Sortieren die Werk- und Briefbände "mischen", die ansonsten nacheinander angezeigt werden.

Natürlich mußte eine sinnvolle Reihenfolge gefunden werden für Texte, die auf den Monat genau datiert sind, solche, denen eine bestimmte Jahreszeit zugeschrieben ist, und diejenigen, für die das Jahr oder sogar nur das ungefähre Jahr feststeht. Die Entscheidung fiel für die Richtung vom Präzisen zum Diffusen.

Wie Gertrud ihre Kinder lehrt	1801 / 13 / 273

Kurztitel — zeitliche Bestimmung | Band | Seite

581. An Graf Karl Johann Christian von Zinzendorf	1783 / B3 / 174

Briefadressat — Briefband

Abb. 6 Beispiel für zwei sinnvolle Kurztitelangaben aus einem Werk- und einem Briefband

Korrekte Hypertext-Anbindungen gewährleisten

Eine zentrale Erleichterung im Verhältnis zur Arbeit mit der Buchausgabe sollte in der zeilengenauen Verknüpfung der eigentlichen Texte Pestalozzis mit den Anhängen 1 und 2 liegen. Während sich im Buch in den Texten keinerlei Hinweise darauf befinden, ob überhaupt Anhanginformationen zur jeweiligen Stelle vorhanden sind, gibt es auf der CD-ROM Hypertextmarken, die durch ihre Position zugleich deutlich machen, in welchem Anhang etwas zu finden ist; außerdem wird der Anhangtext in der oberen Bildschirmzeile angezeigt; ist er länger als eine Zeile, springt der Benutzer über die Marke in den Anhang und ebenso zurück (vgl. Abb. 7: Beispiel für Hypertextmarken neben einem Pestallozzi-Text).

Die Realisierung hätte einfach sein können - dieser Hinweis gilt z. B. Produzenten einer solchen Datenbank, die ihre Buchausgabe zeitgleich zur CD-ROM-Ausgabe vorbereiten -, wenn nicht einige Tücken des Buchsatzes wiederum Mühe bereitet hätten. Außerdem ergaben sich Schwierigkeiten aus Nachlässigkeit bzw. Gedankenlosigkeit beim Scannen, als die CD-ROM-Ausgabe noch nicht konkret anvisiert war. Als sinnvoll erwies sich immerhin die Teilung in Text- und Anhangdateien, die jetzt sozusagen in zwei verschiedenen Datenpools vorlagen.

Fehlermeldungen bei Indexierungsläufen führten auf das Problem der Gültigkeit bzw. Ungültigkeit von Anbindungen: Während in der Buchausgabe die Bezugsstelle "S." links vom Anhangtext abgesetzt ist, war die Einrückung des Textes beim Scannen verlorengegangen. Auf diese Weise waren Literaturhinweise auf bestimmte Seiten in angeführten Schriften an den linken Zeilenrand geraten und wurden aufgrund des "S." als Bezüge zum Pestalozzi-Text interpretiert. Neben absurden Anbindungen führte dies häufig auch zu Abstürzen, wenn eine solche Seite im betreffenden Band beispielsweise nicht existierte. Um diese Fehlerquelle auszuschließen, mußten der komplette Anhangtext kontrolliert und die entsprechenden Zeilenanfänge geändert werden, d. h. die Seitenbezüge auf andere Werke wurden ans Ende der vorausgehenden Zeile gesetzt.

Die Prüfung in bezug auf Richtigkeit der Angaben wurde sowohl durch Satz- als auch durch Einlesefehler notwendig. Typische Zahlenverwechslungen wie 3 und 8 oder 8 und 0 führten natürlich unbemerkt zu falschen Anbindungen. Um diesen Fehlern "auf die Spur zu kommen", ebenso wie übersehenen ungültigen Anbindungen, wurde eine Liste aller Bezüge erstellt, sodann mit einem Programm alle diejenigen Stellen kenntlich gemacht, bei denen eine Ziffer erschien, die niedriger als die vorhergehende war. Es handelte sich - bis auf wenige Reihenfolgen, die bereits im Buch verdreht gesetzt worden waren -, um jeweils eine der beiden Fehlerquellen (vgl. Abb. 8: Beispiel für eine Liste zur Erhebung falscher Reihenfolgen mit Fehlermarkierung).

A2:Die Volksbewegung der Französischen Revolution trug in diesen Tagen

```
696. An Daniel von Fellenberg

▸A Monsieur Fellenberg
 fils de Monsieur le Sénateur
 à Berne.
▸ Neuhof, den 15. September 1792.
 Lieber Freund! Verzeihen Sie, daß ich Ihnen so spät auf Ihr
 freundschaftliches Schreiben antworte. Ich war diese Zeit über
 so zerstreut, daß ich beinahe alles liegen lassen mußte, was nicht
 unumgänglich pressierte.
 Ich danke Ihnen noch mal für die vielen Proben Ihrer
 Freundschaft und Liebe, die Sie mir in hiesigen Gegenden erwiesen, und
 freue mich unendlich, mit Anfang künftigen November ein paar
 Wochen in Ihrem Haus zuzubringen. Ich bitte Sie, Ihrem Herrn
 Papa und Frau Mama für die Güte dieser so freundschaftlichen
 Erlaubnis zu danken.
▸Bis auf diese Zeit wird das Schicksal Frankreichs immer also
▸ entschieden sein, daß man, wenn es unterdrückt [wird], die
▸ Angelegenheit, welche die Menschheit in dem Schicksal dieses Reichs
 hat, freier als jetzo wird beurteilen dürfen, und wenn es bis

PSB 3:282/20 (1792)                                  Gattung: 14.0
F1Info F2        F3        F4        F5Notiz F6Verweis F7↓ F8Zurück F9Suche F10Menü
```

Abb. 7 Beispiel für Hypertextmarken neben einem Pestalozzi-Text

```
V S 191 Z 1/3 die (Armut der) sich schämt
N S 190 Z 4/5 wenn ihr (Gatt)' verhohlt  #10-A

H S 299 Z 37 gekommen (sein mope                #10-A (61)

V S 306 Z 24/25 geschehen? warum sage ich
N S 305 Z 27 an ihr e.s. ist notwendig  einem#10-A

V S 307 Z 40 bis S 308 Z 6 darf den Geist
N S 305 Z 1/2 (nicht) zum Nachteil ihrer K#10-A

V S 308 Z 11 die (Souveranitat> (Hoheit) (
N S 305 Z 11/21 <Nein> (Es ist nicht «ahr#10-A

V S 311 Z 25/26 (Die Unterschrift  nur in
N S 6 Z 5ff  Die Forderungen unter 2 und 3#10-A

V S 68 Z 36 ff  Pestalozzi hat hier mit D
N S 65 Z 28 Konstituenten = Mitglieder de#10-A

H S 193 Zu dem Begriff "Rangieru                #10-A (61)

V S 287 Z 1 ff  Die kontradiktorische Eror
N S 286 Z 29 ff  Das im Fruhjahr 1795 zune#10-A

V S 11 ff  (Der Titel ist von uns hinzuges
N S 3 Z 3 (Der Text folgt) H1 #11-A

V S 6 Z 7 ff  (Der Text folgt) H4
N S 1 Z 32 ff  (Der Text folgt) H5 #11-A

V S 68 Z 34 Gemutsstimmung h2
N S 65 Z 35 richtigen Ebenmaße h2#11-A

H S  Pestalozzis samtliche Werke                #11-A (1))
H S 173-175 liegen daher die Num                #11-A (1))
H S 177-239 um 2 zuruck                         #11-A (1))
H S 240 = Kr  A  256    S 247 =                 #11-A (1))
H S 241 = Kr  A  258        usw                 #11-A (1))
H S 242 = Kr  A  259    S 258 =                 #11-A (1))
H S 243 = Kr  A  260    S 259 =                 #11-A (1))
H S 244 = Kr  A  263    S 261 =                 #11-A (1))
H S 245 = Kr  A  264    S 262 =                 #11-A (1))
H S 246 = Kr  A  265    S 263 =                 #11-A (1))
H S <98> Z 5 Justizbeamter A Jus                #11-A (3))
H S 107 Z 30-S 108 Z 2 12 A = 12                #11-A (9))
H S 108 Z 31-S 109 Z 2 15 A = 15                #11-A (9))
H S 110 Z 30-S 111 Z 5 20 A = 20                #11-A (9))
H S 113 Z 31-S 114 Z 4 26 A = 26                #11-A (9))
H S 114 Z 11-S 116 Z 32 127 A fe                #11-A (9))
H S 116 Z 33-S 117 Z 3 28 A = 27                #11-A (9))
H S 118 Z 24-S 119 Z 2 33 A = 32                #11-A (9))
H S 118 Z 34-S 119 Z 2 bis sich                 #11-A (9))
H S 119 Z 30-S 120 Z 2  35 A = 3                #11-A (9))
H S 120 Z 30-S 121 Z 5 38 A = 37                #11-A (9))
H S 120 Z 33-S 121 Z 1 grodere R                #11-A (9))
H S 121 Z 30-S 122 Z 14 40 A = 3                #11-A (9))
H S 124 Z 21-S 125 Z 2 44 A = 43                #11-A (9))
H S 127 Z 30 - S 128 Z 4 51 A =                 #11-A (9))
```

Abb. 8. Beispiel für eine Liste zur Erhebung falscher Reihenfolgen mit Fehlermarkierung

BEST COPY AVAILABLE

Besondere Bestandteile der beiden Anhänge sind die sogenannten "Kopftexte". Sie enthalten im textkritischen Teil in der Regel Informationen zur Editionsgeschichte, lokalisieren Manuskriptbestände etc.; im sacherklärenden Teil werden Hintergrundinformationen zum Entstehen der Texte geboten. Es schien sinnvoll, diese Kopftexte an die Überschrift des eigentlichen Textes anzubinden, zugleich aber eine Möglichkeit zu schaffen, von jeder beliebigen Textstelle aus an diesen Textanfang springen zu können und nach Einsicht in den Kopftext wieder zurückzugelangen. Wenn beispielsweise ein Nutzer über eine Recherche in einen Text wie "Über Gesetzgebung und Kindermord" gelangt ist und kurz etwas über die Bedeutung des Titels und den Hintergrund der Schrift erfahren möchte, kann er jetzt auf diese Weise verfahren.

Die Kopftexte sind in der Buchausgabe nicht einheitlich gestaltet. Manche tragen die Überschrift der entsprechenden Schrift, manche nicht; einige sind an die erste Seite und Zeile der Schrift angebunden, als handele es sich um einen 1-zu-1-Hypertext. Um diesen Irrtum auszuschließen sowie die nicht durch Seitenbezug kenntlichen Kopftexte als eigenständige Texteinheit zu markieren, die nicht automatisch der jeweils letzten Angabe zum vorherigen Text als Fließtext angehängt wurde, mußten die Kopftexte eine Kodierung erhalten, die sie eindeutig mit einer identischen Kodierung am jeweiligen Anfang des eigentlichen Textes verband. Die nämliche Kodierung erhielt der betreffende Titel im Inhaltsverzeichnis des betreffenden Bandes. Dessen Verbindung mit den Texten stellt ebenfalls eine Hypertext-Anbindung dar.

Im Prinzip wesentlich leichter war die Kopftextmarkierung und -anbindung bei den Briefen dar, die in der Regel jeweils mit der betreffenden Briefnummer beginnen. Allerdings stellte sich auch hier beim Versuch der automatischen Realisierung heraus, daß etliche Einlese- und auch Satzfehler zu Ziffernvertauschungen geführt hatten. Es wurde wiederum über eine Liste aller Kopftext-Anfänge nach unüblichen Reihenfolgen, nach eventuell fehlenden Ziffern und nach Sprüngen gesucht. Die Sprünge betreffen in der Regel Regesten, d. h. der Kopftext hat Bezug zu etlichen, auch nicht explizit angeführten Briefnummern, sondern beginnt mit einer "von ... bis"-Angabe. Auch hier mußte manuell eingegriffen werden, da die Regesten einerseits mehrere Briefe zu einem einzigen Dokument werden ließen - schließlich gilt der Regesten-Text ja für sämtliche angeführte Briefnummern. Wiederum beziehen sich auf dieses eine Dokument mitunter mehrere Kopftexte für einzelne der Briefe (vgl. Abb. 9: Hypertext-Links zwischen den verschiedenen Textebenen).

Hypertext-Links wurden auch zwischen den Registereinträgen und den entsprechenden Fundstellen im Werk gelegt, allerdings nicht eins zu eins, sondern nur zur jeweiligen Seite. Präzise Anbindung hätte nochmals intensive Lektüre und Handarbeit notwendig gemacht, da ja nicht jeweils explizit der

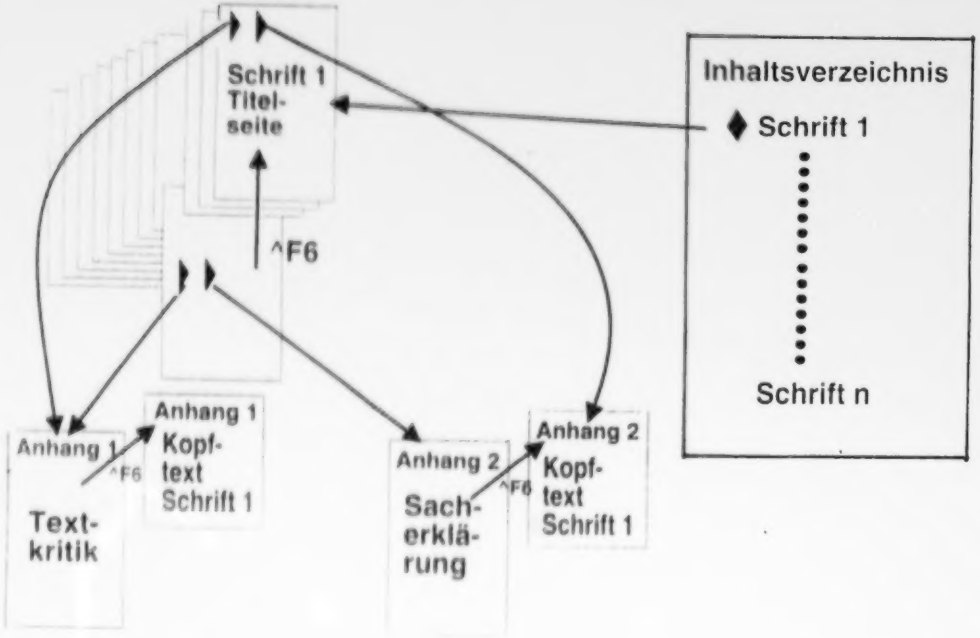


Abb. 9 Hypertext-Links zwischen den verschiedenen Textebenen

Name genannt ist, sondern u.U. eine Variante, nur der Vorname oder eine Umschreibung wie "der Vater" oder "sein Freund".

Systematische Zugriffe ermöglichen

Das Gesamtwerk Pestalozzis weist in der Anlage bereits Strukturen auf, die ein Recherchieren in kleineren Einheiten sinnvoll macht. So wie einzelne Dokumente als Recherchegrundeinheit definiert worden waren, lassen sich auch verschiedene "Typen" von Dokumenten unterscheiden, nämlich die eigentlichen Werke und Briefe einerseits, die Anhänge andererseits. Oft interessieren den Forscher nur Aussagen von Pestalozzi selbst, nicht die - zweifellos wertvollen - Informationen der Anhänge. In anderen Fällen möchten sie lediglich auf eben diese Informationen zugreifen, oder es geht um Auskünfte aus den beigegebenen Registern. Auch der unterschiedliche Umgang mit einem Sachverhalt in offiziellen Texten und in persönlichen Briefen könnte gesondert von Interesse sein. Aus diesem Grunde sind die verschiedenen Dokumententypen separat recherchierbar gemacht.

Unterhalb der Dokumentenebene sind diverse Felder unterschieden, von denen einige - wie einzelne Bände oder Briefe - bereits von der Datenstruktur her automatisch aufrufbar sind. Die Felder "Titel" - also Werk-Überschriften - und "Schlagwort" - bezogen auf die Register und das Glossar - bedurften einer gesonderten Aufbereitung. Überschriften wurden durch spezifische Codes begrenzt, um eine gesonderte Suche nach Titel-Begriffen, z. B. "Recht" oder "Methode", zu ermöglichen. Register-Einträge erfolgten mit einer Begrenzung des eigentlichen Schlagwortes durch eindeutiges, in jedem Eintrag nur einmal verwandtes Satzzeichen.

Die bereits im Hinblick auf die Fundstellenanzeigen vorgenommene eindeutige Adressatenzuschreibung bei Briefen konnte nun automatisch für das Feld "Adressat" in Anspruch genommen werden, das eine Zusammenstellung sämtlicher Briefe an ausgewählte Empfänger ermöglicht. Doch bieten sich auch inhaltlich feinere Unterscheidungen der Texte an, die die Recherche auf systematisch definierte Textmengen einschränkt.

Das Werk Pestalozzis birgt eine Vielzahl von Textgattungen; es enthält neben eigenständigen Abhandlungen sowie Aufsätzen für Zeitschriften und Sammelwerke u.a. zahlreiche öffentliche Aufrufe, Flugblätter, Reden, fiktive Texte und Tagebücher. Dem Wissenschaftler muß sich die Überlegung aufdrängen, daß Aussagen im Rahmen diverser Textsorten verschiedenartig zu gewichten sind bzw. in fiktiven Texten wie den Fabeln mit einer anderen Begrifflichkeit nach Sachverhalten zu recherchieren ist als in einer Abhandlung.

Entsprechend sinnvoll ist die Option, die Recherche auf Dokumente einzelner oder mehrerer Gattungen zu beschränken (vgl. Abb. 10: Übersicht über die Gattungen der Pestalozzi-CD-ROM). Für die Realisierung dieses Vorhabens war es notwendig, eine Aufstellung aller bei Pestalozzi vorkommenden Textgattungen vorzunehmen. Diese ergab sich zum einen aus Anhang-Informationen, zum anderen aus literaturwissenschaftlichen Überlegungen heraus. Den Gattungen wurden bestimmte Codes zugeordnet, die wiederum bei allen betreffenden Texten anzubringen waren. Da es sich jedoch nicht nur um verschiedene Textsorten, sondern innerhalb dieser noch um verschiedene Textzustände handelt, beispielsweise Entwürfe und Fragmente, veröffentlichte und trotz Fertigstellung zu Lebzeiten nicht veröffentlichte Texte, wurde eine Binnendifferenzierung der einzelnen Codes vorgenommen. Diese wiederum erfolgte in bezug auf alle Gattungen gleichsinnig, also alle Entwürfe erhielten z. B. die gleiche Unterziffer, ebenso alle Fragmente usw. (vgl. Abb. 11: Binndendifferenzierung der Gattung "Rede").

Auf diese Weise ist es dem Nutzer der CD-ROM nicht nur möglich, in allen Reden oder nur allen zu Lebzeiten nicht veröffentlichten Reden zu recherchieren, sondern sich etwa sämtliche zu Lebzeiten nicht veröffentlichten Dokumente - von der Flugschrift bis zur Abhandlung - zusammenzustellen und in diesen zu recherchieren, etwa um der wirkungsgeschichtlich relevanten Frage nachzugehen, welche uns bekannten Textbestände den Zeitgenossen Pestalozzis verborgen waren.

Bereits erwähnt wurde die Codierung der Texte im Hinblick auf eine präzise zeitliche Zuordnung der Fundstellenanzeige und eine chronologische Fundstellensortierung. Diese Kennzeichnung ermöglicht zugleich die systematische Einschränkung von Suchanfragen auf bestimmte Zeiträume, z. B. die Periode der Helvetischen Revolution bzw. den Vergleich von Rechercheergebnissen in bezug auf verschiedene Zeitabschnitte. Darüber hinaus läßt sich die Verteilung von Begriffshäufigkeiten auch graphisch als Balkendiagramm präsentieren (vgl. Abb. 12: Präsentation der Verteilung des Begriffs Schule von 1766-1822 (Wirkungszeit Pestalozzis).

Die besondere Relevanz der diversen systematischen Codierungen liegt in deren Kombinierbarkeit. Der Nutzer der CD-ROM kann beispielsweise ermitteln, wie sich Pestalozzi in bestimmten Lebensphasen, etwa nach dem Scheitern seiner Armenanstalt auf dem Neuhof in veröffentlichten und in unveröffentlichten Texten dazu geäußert hat, wie er das Problem der Strafe in Briefen oder in Abhandlungen darstellt, wie er sich nach der Trennung von seinem Mitarbeiter Niederer einerseits in Briefen an ihn, andrerseits in Briefen über ihn geäußert hat (vgl. Abb. 13: Beispiel für eine komplexe Suchanfrage nach dem Begriff Gewalt in allen Entwürfen, Fragmenten und sonstigen zu Lebzeiten Pestalozzis nicht veröffentlichten Texten im Zeitraum 1790 bis Dezember 1799).

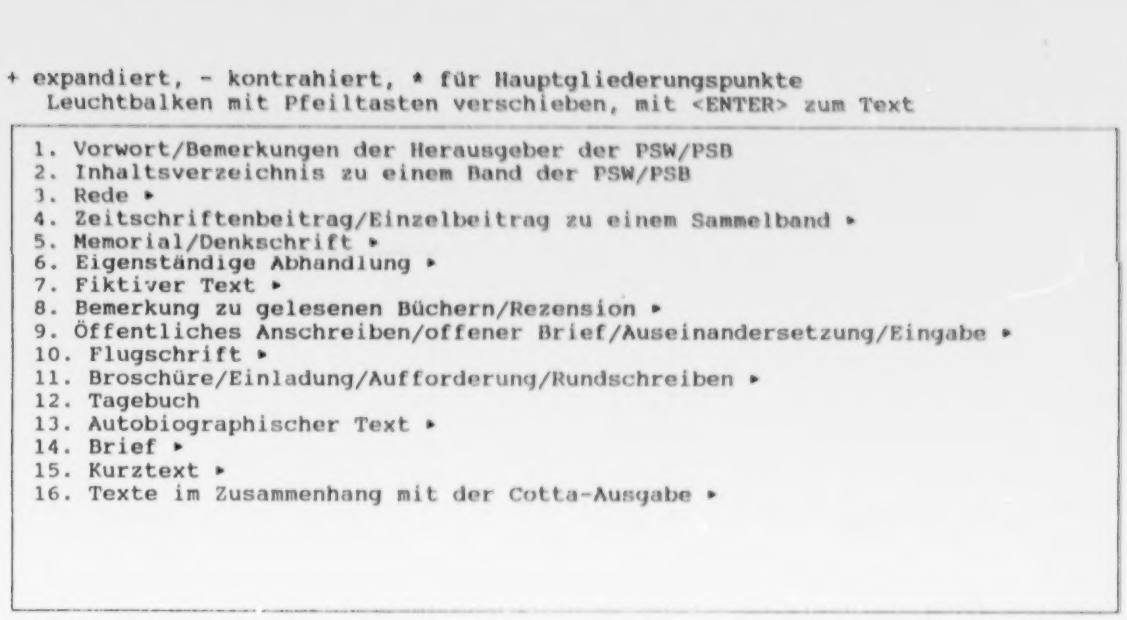


Abb. 10: Übersicht über die Gattungen der Pestalozzi-CD-ROM

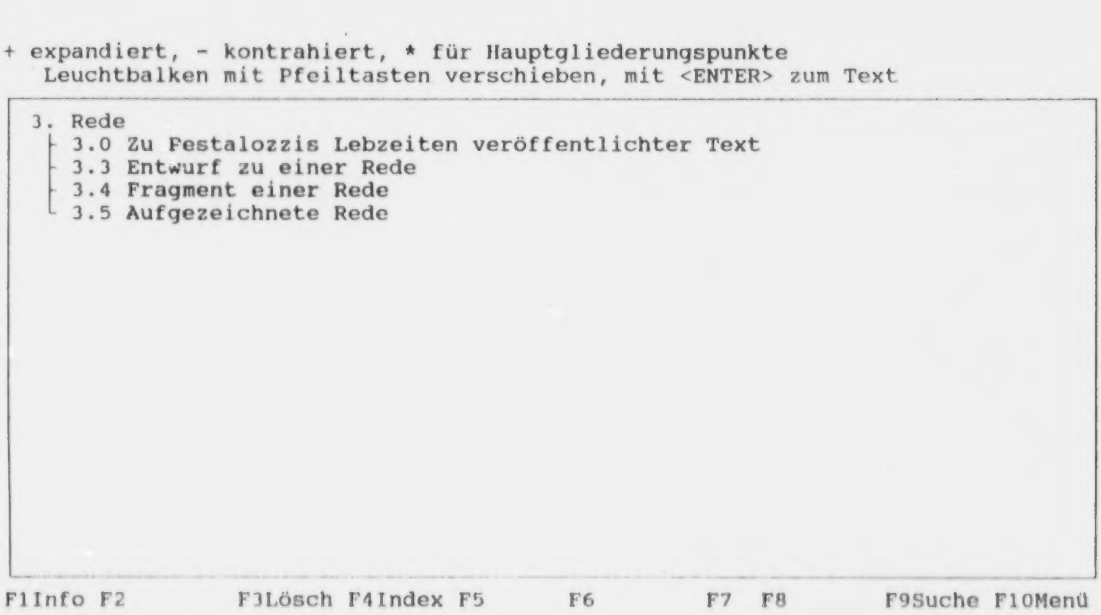


Abb. 11 Binnendifferenzierung der Gattung „Rede"

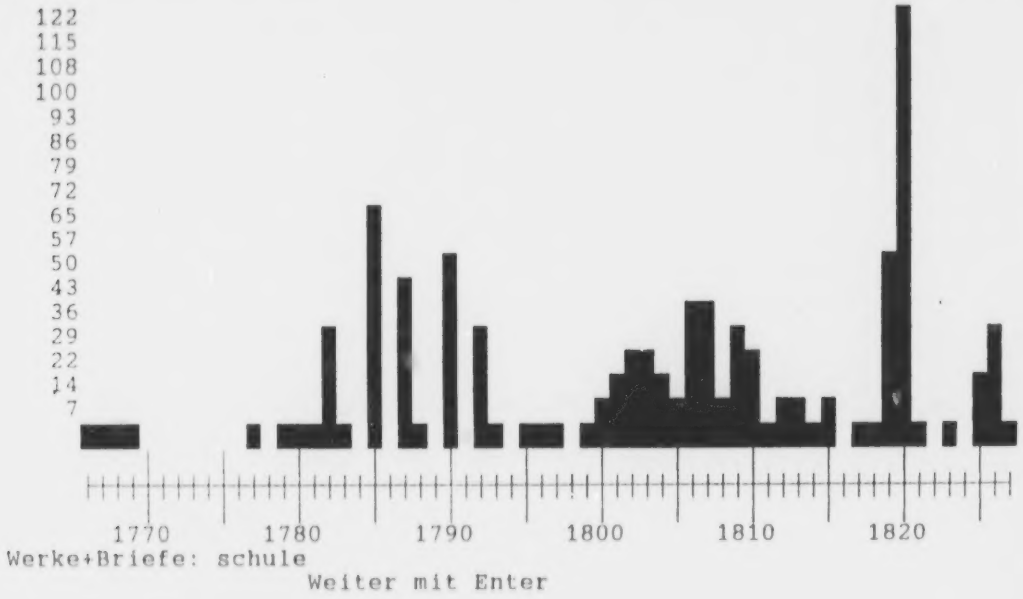


Abb. 12: Präsentation der Verteilung des Begriffs Schule von 1766-1822 (Wirkungszeit Pestalozzis)

Leuchtbalken zum gewünschten Dokument bewegen, Auswahl mit <ENTER>. Blättern F7

SUCHE IN: Werke und Briefe Von: 01/90 Bis: 12/99

(Gattung:*.3 oder Gattung:*.4 oder Gattung:*.5) Gewalt

Suche nach: (Gattung:*.3 oder Gattung:*.4 oder Gattun	Ergebnis: 56
↑ 40 Zu Condorcets Esquisse d'un tableau historique	1797/11/39
41 Oratio pro Domo	1797/11/79
42 Eine Erklärung zu dieser Antwort	1797/11/341
43 Das Unglück des Landes	1797/11/347
44 Mensch und Hund	1797/11/351
45 Was das Gedeihen fördert	1797/11/353
46 Fragment	1797/11/359
47 Predigt an die Franzosen	1797/11/43
↓ 48 An die Freunde der Freiheit am Zürichsee und der Enden	1797/10/303

F1Info F2 F3 F4 F5 F6 F7↓ F8 F9Suche F10Menü

Abb. 13: Beispiel für eine komplexe Suchanfrage nach dem Begriff Gewalt in allen Entwürfen, Fragmenten und sonstigen zu Lebzeiten Pestalozzis nicht veröffentlichten Texten im Zeitraum Januar 1790 bis Dezember 1799

Ergänzende Hinweise anbieten

Zuzüglich zur bislang erschienenen Pestalozzi-Ausgabe wurden der CD-ROM, wie bereits mehrfach angesprochen, der noch im Druck befindliche erste Registerband sowie ein für den zweiten Registerband vorgesehenes Glossar beigegeben.

Der Registerband enthält diverse Gesamtnamenregister, vor allem das als Kurzlexikon konzipierte Verzeichnis der Personennamen, u.a. auch separat zusammengestellte Register der Empfänger von Pestalozzis Briefen und der Subskribenten der ersten Pestalozzi-Gesamtausgabe bei Cotta.

Ein besonderer Vorteil für den CD-ROM-Nutzer im Gegensatz zum Leser des Register-Bandes erwächst aus der Tatsache, daß die Autoren Friedrich/Springer bei Einträgen zu verschiedenen Personen, die jeweils den gleichen Sachverhalt beschreiben, beispielsweise "Schüler" oder "Lehrer in Pestalozzis Anstalt in Burgdorf", jeweils die gleiche Formulierung gewählt haben.

Auf diese Weise läßt sich bei einer Beschränkung auf den Dokumententyp "Personennamenregister" beispielsweise eine Zusammenstellung aller von Pestalozzi bestellten Bücher - vermerkt bei den betreffenden Autoren - erzeugen. Eine Eingrenzung auf den Typ "Briefempfänger" oder "Subskribenten der Cotta-Ausgabe" ermöglicht eine Zusammenstellung aller Händler und Lehrer oder aller Wiesbadener unter diesen. Der bereits für die Buchausgabe vorgenommenen systematischen Gruppierung der Personenkreise nach geographischer Herkunft, Berufsgruppen und gesellschaftlichen bzw. politischen Funktionen kommt im Hinblick auf elektronische Recherchemöglichkeiten nun eine besonders nutzerfreundliche Bedeutung zu.

Das Glossar - im geplanten zweiten Registerband auf die Buchausgabe bezogen - enthält Begriffe in zahlreichen Schreibvarianten, die auf der CD-ROM bereits nivelliert wurden. Es bedurfte darum einer gesonderten Durchsicht im Hinblick auf das Vorkommen der fraglichen Schreibweisen bzw. der Begriffe auf der CD überhaupt.

Eine weitere wichtige Informationszugabe besteht in einer Zusammenstellung von Abbildungen. Es handelt sich dabei u.a. um eine Vielzahl von Pestalozzi-Portraits, um Darstellungen von Familienangehörigen, Freunden und Gegnern sowie Mitarbeitern und Schülern, um Abbildungen von Wirkungsstätten, von Lehrmitteln Pestalozzis wie diversen Anschauungstafeln und von Faksimiles von Handschriften. Diese Bilder erfüllen vor allem die Funktion der Veranschaulichung und sind zu diesem Zweck an entsprechende Textstellen mit Hypertext-Marken angebunden, im wesentlichen bei den Namen-

registern, darüber hinaus bei besonders klaren Textbezügen. Sie lassen sich aber auch gesondert als Dokumententyp "Abbildungen" aufrufen und aus thematisch gruppierten Inhaltsverzeichnissen heraus anwählen. Um die Bilder in dieser Weise zu präsentieren und zu verknüpfen, waren zum einen eine Gruppierung und das Verfassen von hinreichenden Bilderläuterungen, zum anderen das Setzen von Codes an den entsprechenden Textstellen erforderlich.

Der zur Veranschaulichung des Textbearbeitungsgrades in Images beigegebene Werkband 1 wurde - nach eindeutiger Benennung der Bilddateien - den jeweiligen Seitenanfängen automatisch zugeordnet.

Zum Schluß ...

Die Aufbereitung und systematische Strukturierung der Pestalozzi-Volltextdatenbank erfolgte aus dem Gesichtswinkel der Wissenschaft. Es waren konkrete Forschungsanliegen, die die Buchausgabe nicht oder nur schwerlich zu befriedigen geeignet ist, die zum Träumen von einer CD-ROM, zur Formulierung eines Pflichtenheftes, zur Auswahl geeigneter Software und zur Bearbeitung des Volltextes führten.

Im Zuge dieser Arbeit erfolgten fortwährend Impulse zu weiteren Möglichkeiten des gezielten, systematischen Recherchierens, zu verschiedenartigen Optionen der Ablage und Weiterverarbeitung usw.

So unendlich wie das "Meer der Fragen" in der Wissenschaft, so unendlich ist auch die Anzahl möglicher geeigneter "Fanggeräte" und "Fangtechniken". Neue Projekte für neue "Ozeane" drängen sich bereits auf.

Anschrift der Verfasserin
Sylvia Springer
Universität Düsseldorf
FB Erziehungswissenschaften
Universitätsstr. 1
40225 Düsseldorf
Tel.: (02 11) 311 36 80

Doris Bambey

Vorbereitung und Entwicklung der CD-ROM Bildung

Probleme, Methoden und Instrumente

1 Die Ausgangslage

Die Literaturdokumentation Bildung auf CD-ROM in der ersten Fassung von 1994 bietet den Zugriff auf rund 200 000 Dokumente aus den Jahren 1980 bis 1992. Für das jährlich geplante Update werden jeweils etwa 30 000-35 000 zusätzliche Literaturnachweise zu erwarten sein. Die der CD-ROM zugrundeliegende Datenbasis setzt sich zusammen aus Zulieferungen von 12 Instituten (in Zukunft noch mehr). Diese Institute, die teilweise schon 15 bis 20 Jahre ihre speziellen Teilbereiche des Themenfeldes Bildung dokumentieren, repräsentieren ein breites Spektrum an Erfassungs- und Verschlagwortungspraxen.

Dies bedingt, daß die gelieferten Datenformate von äußerst heterogener Struktur sind: Bei einer größeren Gruppe der Zulieferer wird die Datenerfassung nach den Regeln des ehemaligen DOPAED (Dokumentationsring Pädagogik) vorgenommen, oftmals sind diese Regeln jedoch im Laufe der Jahre nach den eigenen Bedürfnissen und technischen Möglichkeiten modifiziert worden. Institute aus dem bibliothekarischen Bereich orientieren sich wiederum an Regeln, die auf den Austausch innerhalb von Verbundsystemen abgestimmt sind (z. B. MAB). Einzelne Institute schließlich erfassen ihre Daten weitgehend den gewachsenen internen Anforderungen entsprechend.

Dieselbe Ausgangsproblematik betrifft auch die Verschlagwortungspraxis. Neben dem (zum Teil veralteten) *Thesaurus Pädagogik* werden verschiedene Subthesauri verwandt, es wird zudem auch nach der Schlagwortnormdatei vorgegangen oder auf der Grundlage von - mehr oder weniger kontrollierten - eigenen Schlagwortlisten indexiert. Ein Teil der Institute bevorzugt die Verschlagwortung mittels zerlegter Begriffe (postkoordiniertes Vokabular), andere wiederum indexieren bevorzugt mit Mehrwortgruppen und Komposita (präkoordiniertes Vokabular) oder bilden Schlagwortketten.

Die Frage ist, mit welchen Werkzeugen und in welchen Arbeitsschritten man aus dieser disparaten Datenbasis einen Bestand entwickeln kann, der den Anforderungen einer elektronischen Publikation - hier CD-ROM - genügt.

Vier Bearbeitungschritte sind hierbei von besonderer Relevanz:

1. Konvertierung der zugelieferten Ausgangsformate in ein einheitliches Arbeitsformat,
2. Vereinheitlichung der Feldinhalte, speziell bei den recherche- und prüfrelevanten Feldern,
3. effektive und qualitativ orientierte Dublettenkontrolle,
4. Normierung und Kontrolle der Verschlagwortung.

2 Erstellung eines Arbeitsformates, Normierung von Feldinhalten

2.1 Kriterien für ein Arbeitsformat

Zusammengesetzte Felder sind nach Möglichkeit zu vermeiden; denn eine analytische Struktur der Daten ermöglicht:

- die zielgenaue Korrektur der Feldwerte aufgrund von präzisen, nach Typen getrennten Indexlisten. Hierdurch steigert sich die Transparenz der Datenbank,
- die Sortierung der Datensätze nach jedem gewünschten Kriterium,
- flexible Generierung von Such- und Ausgabemasken mit den Feldern, die gerade gewünscht werden,
- flexible Ausgangsbasis für weitere Datenaustauschprozesse.

Obligatorische Felder sollten auch de facto belegt sein, um die Recherche kalkulierbarer zu machen und einen Mindeststandard an Vollständigkeit der Informationen pro Datensatz zu gewährleisten. (Z. B. Sprachangaben, Literaturtyp). Vorrangig verwendete Such- und Prüffelder müssen von der Ansetzung und der Rechtschreibung her in möglichst einwandfreiem Zustand sein. Nur so ist eine effektive Dublettenkontrolle zu erzielen.

Auch die Recherchierenden freuen sich später, wenn sie ihren Autor nicht mühselig einmal unter "Kleist, Heinrich von" (RAK) und zudem unter "Von Kleist, Heinrich" (Nicht RAK) suchen müssen.

2.2. Probleme bei der Datenkonvertierung

Schwierigkeiten bei der Konvertierung von zugelieferten Daten treten zunehmend dann auf, wenn die Datenstruktur synthetisch aufgebaut ist, d. h. Informationselemente unterschiedlichen Typs - schlimmstenfalls ohne eindeutige Trenner - in einem Feld untergebracht sind. Ist die Datenstruktur dahingegen analytisch aufgebaut, also jedes Informationselement eines Typs seperat in einem Feld abgelegt, vereinfacht dies die Konvertierung grundlegend.

Als Beispiel für ein z. T. synthetisches Format mit verschiedenen Schwierigkeitsgraden kann das DOPAED-Format dienen: Das Feld "Erscheinungsvermerk" ist zusammengesetzt aus mehreren unterschiedlichen Informationen, die jedoch eindeutige Trenner aufweisen: Erscheinungsvermerk: Ort: Verlag (Jahr), Seiten. Hier muß bei der Konvertierung lediglich die Möglichkeit berücksichtigt werden, daß ein oder mehrere Informationsteile fehlen. Das Titelfeld bringt gravierendere Probleme mit sich: /Hauptsachtitel/. Zusatz zum Hauptsachtitel. Auflage.

Der Zusatz zum Hauptsachtitel und die Auflagebezeichnung sind lediglich durch einen Punkt getrennt. Punkte können aber schon Bestandteil des Hauptsachtitels sein, z. B.: "... und Anton" oder "Das Wirken Dr. Schweizers". In diesem Fall ist eine korrekte Zerlegung der zu identifizierenden Informationselemente nicht gewährleistet.

Die präzise und sichere Idendifikation der Auflagebezeichnung z. B. ist jedoch von entscheidender Bedeutung für die Dublettenprüfung. Wird bei der Prüfung die Auflagezeichnug nicht in den Prüfschlüssel übernommen, werden die betreffenden Dokumente irrtümlich als Dubletten qualifiziert.

2.3 Programme

Drei Programme spielen bei diesen Konvertierungs- und Normierungsarbeiten eine wesentliche Rolle:

- das Konvertierungsprogramm Infotrans,
- das Textverarbeitungssystem Wordperfect, speziell der Arbeitsbereich Macros,
- das Information-Retrievalsystem LARS.

Infotrans arbeitet auf PC-Basis und funktioniert auf einem relativ abstrakten Niveau mittels Variablen. D. h., es müssen nicht wie bei den Makros der Textverarbeitungssysteme konkrete Einzelnennung des Umzusetzenden angegeben werden (setze a um in A, setze b um in B), sondern es sind vielmehr

Klassen von Umsetzungen definierbar (setze alle Kleinbuchstaben um in Großbuchstaben oder markiere alle Ziffern in den ersten drei Wörtern des Hauptsachtitels). Diese Leistung ist die Voraussetzung, um komplexe Konvertierungsfälle überhaupt zu bewältigen.

Infotrans arbeitet sowohl auf Feldebene als auch global auf Dateiebene. Die Suche-und-Ersetze-Befehle werden in Umsetzungstabellen abgelegt. Das Programm ist von seinen Anforderungen her auf die geübteren Datenbankbearbeiter zugeschnitten, die jedoch keine eigenen programmiertechnischen Vorkenntnisse besitzen müssen.

LARS (hier für DOS) ermöglicht die Korrektur von Feldinhalten über Indices, d. h., die Suche-Ersetze-Befehle, die innerhalb des Menüs "Index" definiert werden, werden automatisch - und somit schnell und einheitlich - in den betreffenden Dokumenten ausgeführt. Bei Feldern, die nicht als Indexfelder definiert oder deren Änderungen zu komplex sind, besteht die Möglichkeit, mit dem sog. "Änderungsdienst" Teile der Dokumente auszugeben (z. B. nur Primärschlüssel und Titel), außerhalb von LARS dann mittels Infotrans oder Textverarbeitung zu korrigieren und anschließend wieder an die korrekte Stelle des Ausgangsdokuments zurückzuspielen.

3 Effektive und qualitativ orientierte Dublettenprüfung

Die vorhergehenden Arbeiten an den Datenbeständen sind die Voraussetzung für eine möglichst effektive Dublettenprüfung. D. h., wenn der aus bestimmten Feldinhalten oder Teilen von Feldinhalten (z. B. Nachname des Autors, drei erste Wörter des Titels, Jahrgang) zusammengesetzte Prüfschlüssel korrekt funktionieren soll, müssen auch die verwendeten Feldwerte korrekt und einheitlich vorliegen. Nicht nur der quantitative Aspekt (möglichst alle Dubletten finden) sollte bei der Prüfroutine eine Rolle spielen, sondern auch qualitative, die Datengüte steigernde Momente. Das Problem besteht darin, daß bei herkömmlichen Prüfverfahren wertvolle Informationen durch die generelle Eliminierung der Dubletten ungenutzt bleiben. Bei einer erfahrungsgemäß bis zu 30 % betragenden Dublettenrate wird der Verlust von Informationen durch ein mechanisches "ja-nein-Verfahren" gravierend.

Programm

Bedingung für das auf diese Problemstellung abgestimmte Dubletten-Prüfprogramm (auf Clipper-Basis) ist, daß es nicht nur auf der Grundlage der "ja-nein-Entscheidung" arbeitet, sondern auch die Übernahme von Teilen abgelehnter Dokumente in das ausgewählte Dokument ermöglicht.

Beispiel:

DOKUMENT 1

- Bibliographische Angaben: korrekt und vollständig,
- Standortangabe für Fernleihe: enthalten,
- Beschlagwortung: sehr grob und lediglich für Bibliothekskatalog,
- Abstract: nicht vorhanden.

DOKUMENT 2

- bibliographische Angaben: weisen Fehler auf,
- Beschlagwortung: sehr umfangreich und präzise,
- Abstract: enthalten.

DOKUMENT 3

- bibliographische Angaben: weisen Fehler auf,
- Beschlagwortung: grob,
- Abstract: nicht enthalten,
- Notation der DOPAED-Fachsystematik: enthalten.

Das Dokument, das in die Datenbank übernommen wird, enthält letztendlich die Vorzüge aller abgelehnten Dokumente:

ZIELDOKUMENT

- Korrekte bibliographische Angaben inkl. Standortangabe für die Fernleihe,
- ausführliche, fachgerechte Verschlagwortung,
- Abstract,
- Systematikstelle zur Erstellung der *Bibliographie Pädagogik*.

Das Verfahren des "composed documents" dauert länger, da es nicht als Routineablauf automatisch durchführbar ist; es bedarf vielmehr des intellektuellen Abgleichs der Dokumente. Dafür erreicht man eine Veredelung des Dokuments mit vergleichsweise einfachen Mitteln. Die urheberrechtliche Problematik schließt sich hier natürlich als Problempunkt an - sie will ich hier nicht näher erläutern. Außerdem sollte man sehr vorsichtig sein mit dem Mixen verschiedener Indexate, also dem Addieren von Schlagwörtern verschie-

dener Institute, da die Gefahr besteht, daß der Informationsgehalt beliebig wird und die Stimmigkeit in sich verloren geht. Das Zusammenführen von Indexaten wird von uns bisher nur ansatzweise und gezielt durchgeführt, hierzu besteht noch großer Diskussionsbedarf.

4 Kontrolle und Normierung der Verschlagwortung

Aufgrund der Verwendung verschiedenster Subthesauri und Schlagwort- bzw. Stichwortlisten hätte man ohne Normierung des Zuliefererervokabulars mit einem Schlagwortaufkommen von über 100.000 bis 130.000 Einzeltermen rechnen müssen. Die Erarbeitung eines aktuellen "Thesaurus Bildung" als bindende Indexierungs-und Recherchegrundlage wird weder als pragmatisch eingestuft noch als den gewachsenen und spezifischen Anforderungen der Einzelinstitute adäquat.

Die pragmatische Lösung geht in die Richtung, das vorhandene Vokabular derart zu bearbeiten, daß folgende Probleme berücksichtigt sind:

- Normierung von Schreibvarianten (Naturwissenschaft--> Naturwissenschaften),
- Homonym- bzw. Polysemkontrolle (Bewegung--> (Bewegung (Pol), Bewegung (motorische), Bewegung (Phys)),
- Synonymkontrolle (Andragogik <--> Erwachsenenbildung),
- Zerlegung von zu komplexen Wortgruppen oder Komposita, (Berufsschulreligionslehrer --> Berufsschule, Religionslehrer).

Eine wesentliche Bedingung besteht darin, daß die Synonyme als gleichrangige definiert sind, d. h., bei der Arbeit mit der CD-ROM soll es für die Recherchierenden egal sein, ob sie mit dem Schlagwort "Andragogik" oder mit dem Schlagwort "Erwachsenenbildung" suchen. Die Retrieval-Software CD-Answer garantiert in jedem Fall, durch Zusammenführung der Synonyme im Hintergrund, daß im Ergebnis alle Dokumente aufgeführt sind, die sowohl das eine als auch das andere Schlagwort enthalten. Die Strukturierung der Suchpfade bewirkt somit eine Erhöhung der Zugangsmöglichkeiten, beläßt das Vokabular in seiner fachlich gebotenen Spezifik und garantiert gleichzeitig - durch die Synonymzusammenführung - eine kontrollierte Recherche.

Programme

Drei Programme kommen hauptsächlich zum Einsatz:

- das Thesaurusverwaltungsprogramm Proterm,
- ein Schlagwortkonvertierungsprogramm von PROGRIS,
- ein Schlagwort-check-Programm (Clipper-Basis).

Mittels eines Checkprogrammes, welches sowohl Schlagwortdubletten als auch neue Werte identifiziert und in gesonderte Dateien ablegt, werden die jeweils noch nicht in Proterm enthaltenen Schlagwörter ermittelt und zu Importlisten zusammengestellt. Jedes Schlagwort in Proterm wird mittels Notation einer Systematikstelle der DOPAED-Fachsystematik zugeordnet. Die Notation ermöglichen die Zusammenstellung der Schlagwörter nach Fachgruppen, innerhalb derer sowohl Synonyme als auch nicht zugelassene Schreibvarianten ermittelt werden.

Diese Struktur wird in ein Schlagwortkonvertierungsprogramm übertragen. Es ist in der Lage:

- innerhalb des Dokumentes auf das Deskriptorenfeld zuzugreifen,
- die Schlagworte umgesetzt in ein Zielfeld zu schreiben,
- nicht identifizierbare Schlagwörter in ein Kontrollfeld zu übertragen,
- eine Schlagwortdublettenkontrolle durchzuführen (wichtig bei Schlagwortketten z. B. Bildung:Recht, Bildung:Internationaler Vergleich).

Die Schlagwörter, die sich im Kontrollfeld befinden, sind überwiegend Terme, die in einfachere Eineiten zerlegt werden müssen oder als Homonyme zu definieren sind, letzteres kann (leider) nur manuell unter Sichtung der jeweiligen Dokumente vorgenommen werden. Übrig bleiben auf der CD-ROM letztendlich noch ca. 25 000 Schlagwörter, wovon ungefähr 10% in Synonymrelationen eingebunden sind.

Anschrift der Verfasserin:

Doris Bambey
Modellversuch FIS Bildung
Berliner Str. 31-35
65750 Eschborn
Tel.: (06196) 48 41-45; Fax: -61

Alexander Botte

Workshop 1: "Erfahrungen in der Produktion von elektronischen Diensten"

Einleitende Bemerkungen und Zusammenfassung

Hintergrund dieses Workshops war der in der einschlägigen Fachpresse verfolgbare weltweite Trend zu steigenden Nutzerzahlen bei Informationsangeboten, die auf dem Medium CD-ROM aufsetzen. Ablesbar ist diese Entwicklung unter anderem am exponential wachsenden Absatz von CD-ROM-Laufwerken, auch im privaten Sektor.

Parallel dazu findet eine rasante Entwicklung zur Vereinfachung der rein technischen Abläufe statt, die begleitet ist von einer steigenden Zahl von Anbietern von Dienstleistungen auf diesem Gebiet. Damit eröffnen sich neue Möglichkeiten, auf relativ einfache Weise, CD-ROM-Publikationen in Eigenregie zu erstellen. Eine wichtige Folge davon ist, daß der Prozeß der Herstellung von CD-ROMs und die Preisentwicklung zwischenzeitlich auch und gerade für kleinere Produktionsmengen das Medium für Hersteller attraktiv machen, die nicht auf große verlegerische Erfahrungen und Ressourcen zurückgreifen können.

Diese Möglichkeiten waren Anlaß des Workshop-Themas: CD-ROM-Produktionen standen also im Mittelpunkt.

In drei Erfahrungsberichten aus der Praxis, die aber auch allgemeines theoretisches Hintergrundwissen zu vermitteln wußten, wurden unterschiedliche Aspekte der CD-ROM-Produktion beleuchtet. Gegenstand waren also nicht die inhaltlichen Aspekte, sondern die organisatorischen und technischen Voraussetzungen und Abläufe. Konkrete Zielsetzung war es, Orientierung zu geben für die Veröffentlichung eines Informationsproduktes auf CD-ROM.

Der Charakter der Veranstaltung mit durchweg sehr dichten Vorträgen ließ leider die Ausgestaltung zu workshopähnlichen Elementen nicht zu. Der enge Zeitrahmen erlaubte nach den Vorträgen lediglich die Beantwortung von Verständnisfragen; praktische Demonstrationen und die Diskussion von Einzelfallanwendungen mußten aufgrund der begrenzten Zeit entfallen.

Anschrift des Verfassers

Alexander Botte
Leiter Modellversuch FIS Bildung
Berliner Str. 31-35
65760 Eschborn
Tel.: (0 61 96) 4841-45; Fax: -61

Heinz Ziegler

Einige Anforderungen an Recherchesoftware bei Verwendung eines Thesaurus

0 Einführung

Nach wie vor arbeiten die meisten Literaturdatenbanken (Referenzdatenbanken) mit einem Schlagwortfeld, dem ein kontrolliertes Vokabular zugrunde liegt. Fast immer sind die Schlagwortlisten strukturiert: In einfachen Fällen ist das Vokabular in einige Hauptgruppen gegliedert, und es gibt Synonymverweise; in vielen Fällen aber sind die Strukturen differenzierter in Richtung eines Thesaurus ausgebaut.

Auch die im Aufbau befindliche Literaturdatenbank Berufliche Bildung[1] wird für das Schlagwortfeld ein kontrolliertes und strukturiertes Wörterbuch benutzen, weil nach unserer Auffassung auf diese Weise der Benutzer, und insbesondere der gelegentliche Benutzer, noch am ehesten zu den gesuchten Informationen findet. Eine Voraussetzung hierfür ist freilich eine leistungsfähige Recherchesoftware mit einer gut selbsterklärenden und einfach zu bedienenden Oberfläche, eine Software, die dem Benutzer die durch die Strukturierung und vielfache Verknüpfung des Wortguts gebotenen Möglichkeiten zur Hinführung auf für ihn relevante Suchwörter sowie zur Erleichterung und Verbesserung der Recherche erschließt. Erstaunlicherweise haben wir eine solche Software - für PC, möglichst mit Fenstertechnik und Mausbedienung, netzfähig und zu einem erschwinglichen Preis - bisher nicht gefunden. Zwar gibt es alle dafür erforderlichen Komponenten in voller Ausprägung in dem einen oder anderen Produkt, aber bisher hat offenbar noch kein Softwareher-

[1] Näheres zu dem von Prof. Dr. *Peter Diepold*, Humboldt-Universität zu Berlin, geleiteten Projekt siehe: Literaturdokumentation Berufliche Bildung : Ziele und Fortschritte eines vom BMBW geförderten Projekts / *Heinz Ziegler*. - In: Bildungsdokumentation heute : Fachinformation für Wissenschaft und Praxis ; 1. GIB-Fachtagung Frankfurt/Main, 10.-11. November 1993 / *Peter Diepold, Diann Rusch-Feja* (Hrsg.). Berlin : Gesellschaft Information Bildung, 1994. S. 39-48. (Schriftenreihe der Gesellschaft Information Bildung ; 1).

steller diese Komponenten in einem Programm vereinigt. Vielleicht macht dieser Erfahrungsaustausch auf ein solches Produkt aufmerksam.

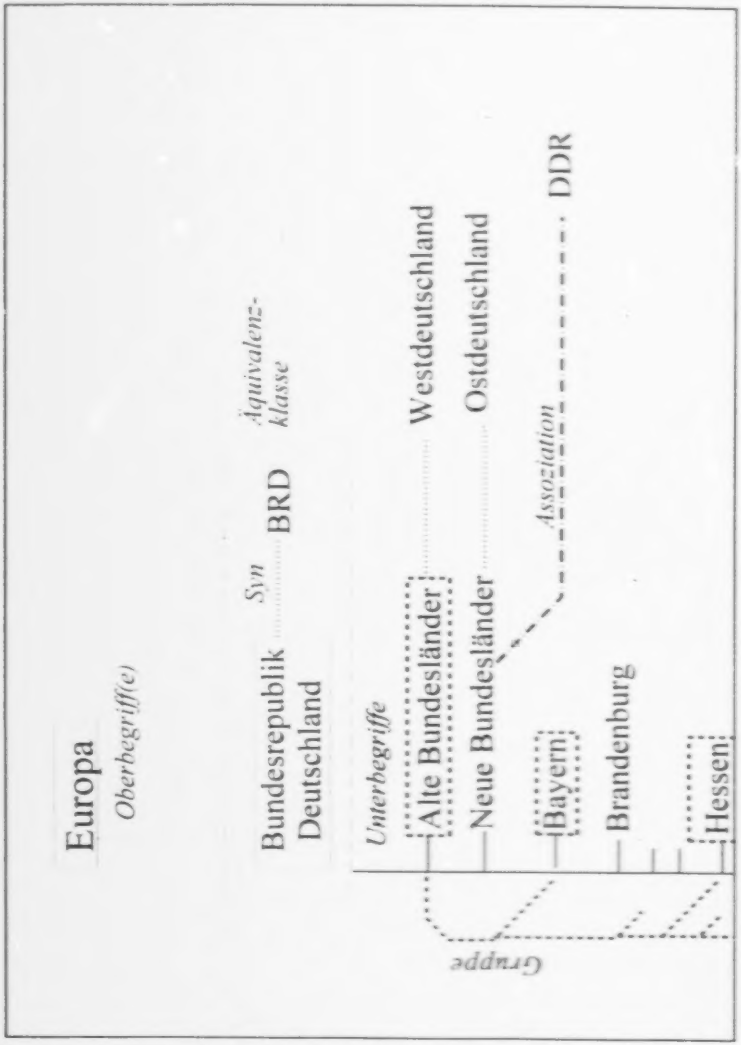
Die Entscheidung für eine bestimmte Software hat auch Folgen für den Datenbasenhersteller, vor allem hinsichtlich der Indexierung. Läßt es beispielsweise die Recherchesoftware nicht zu, bei Eingabe des allgemeinen Suchbegriffs HOCHSCHULE wahlweise zugleich automatisch nach allen spezifischeren Unterbegriffen - wie UNIVERSITÄT, FACHHOCHSCHULE, KUNSTHOCHSCHULE usw. zu suchen, so muß man entweder dem Rechercheur zumuten, alle diese Begriffe einzeln einzugeben (was gewiß nicht nutzerfreundlich wäre), oder es muß beim Indexieren des Dokuments beispielsweise zu dem spezifischen Begriff FACHHOCHSCHULE zusätzlich der Oberbegriff HOCHSCHULE vergeben werden. Bereits dieses einfache Beispiel zeigt die Rückwirkung der Recherchesoftware auf die Gestaltung der Datenbank.

Wegen dieser Zusammenhänge wurden für die Literaturdatenbank Berufliche Bildung "**Anforderungen an die Recherchesoftware**" zusammengestellt. Im folgenden werden hieraus einige auf den Thesaurus bezogene Anforderungen betrachtet. Das ist freilich nur einer der für die Bewertung wichtigen Aspekte; aber in Anbetracht des Aufwandes, den die Erarbeitung und Pflege eines strukturierten Wörterbuchs erfordert, ist die Hervorhebung der Frage schon berechtigt, welche Funktionen die Software anbieten sollte, damit die durch ein strukturiertes Wörterbuch gegebenen Möglichkeiten leichteren und erfolgreicheren Recherchierens auch genutzt werden können.

1 Anforderungen

1.0 Vorbemerkung

Die verschiedenen Software-Handbücher und Anleitungen benutzen für bestimmte Strukturbeziehungen unterschiedliche Bezeichnungen. Im vorliegenden Aufsatz wird wie folgt verfahren: Als **Synonyme** werden alle Begriffe bezeichnet, die der gleichen Synonymgruppe, der gleichen (bedingten) Äquivalenzklasse zugeordnet sind, unabhängig davon, ob es sich um echte Synonyme handelt. Der Begriff **Assoziation** oder assoziative Beziehung wird im Sinne eines "Siehe-auch-Verweises" gebraucht; es ist der Hinweis von einem Begriff auf möglicherweise relevante andere Begriffe, die (meist) in anderen Hierarchiezweigen stehen (zum Beispiel der Hinweis von EUROPÄISCHE UNION als politische Organisation auf EU-LÄNDER als Ländergruppe oder von PERSONALPLANUNG auf PERSONALABBAU). *Bild 1* und *Bild 2* veranschaulichen diese Beziehungen.



139

Bild 1 Hierarchie (Ober-/Unterbegriffe), Äquivalenz (Synonyme), Assoziation (Siehe-auch-Verweis)

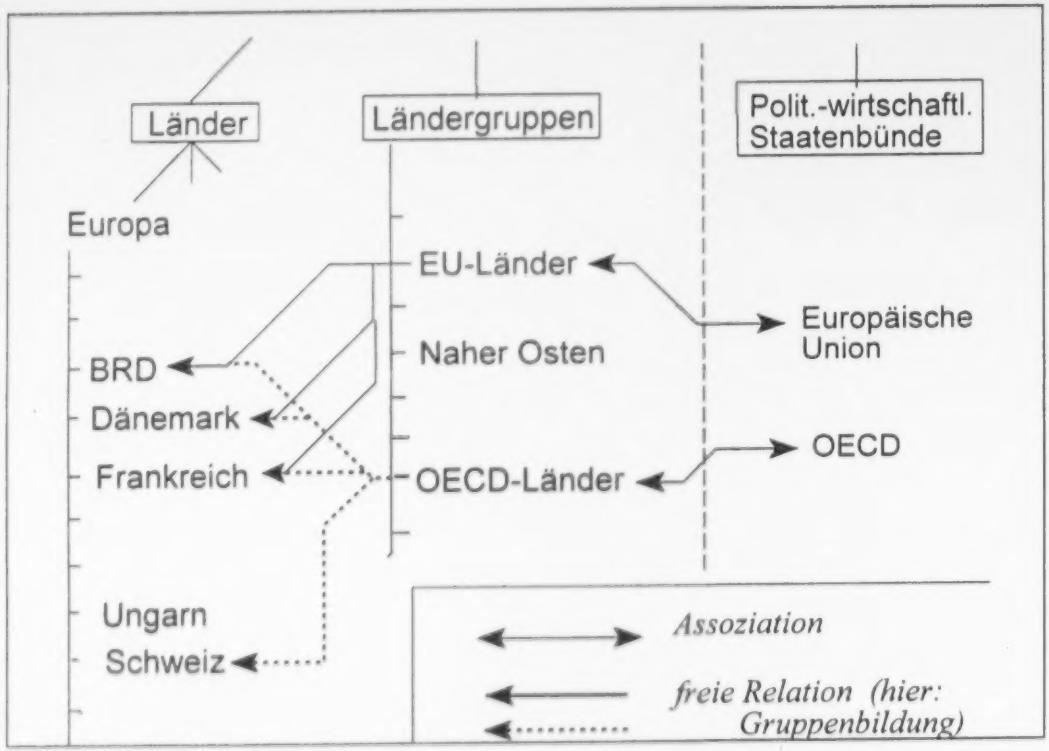


Bild 2 Gruppenbildung mittels freier Relation: Assoziation

1.1 Verwaltete Strukturen

- **Verwaltung**. Im strukturierten Wörterbuch sollen mindestens folgende Beziehungsarten verwaltet werden:
  - *Synonymie* (bedingte Äquivalenzklassen)
  - *Hierarchie* mit der Möglichkeit der *Polyhierarchie*
  - *Assoziation*
- **Gruppenbildung.** Erwünscht ist ferner die Verfügbarkeit von mindestens einer weiteren, *frei definierbaren Relation* (für Gruppenbildung ohne Beachtung der Primär-Hierarchie).
- **"Freie" Schlagwörter**. Im strukturierten Wörterbuch sollen auch Begriffe verwaltet werden, die (noch) nicht in die Struktur eingeordnet sind.
- **Textliche Erläuterung.** Es soll möglich sein, einem Begriff eine kurze textliche Erläuterung zuzuordnen.
- **Anzahl der Relationen.** Die Anzahl der von einem Begriff ausgehenden Beziehungen zu anderen Begriffen soll nicht begrenzt sein.
- **Länge.** Die zulässige Länge für einen Wörterbuchbegriff soll nicht kleiner als etwa 100 Zeichen sein.

1.2 Anzeigemöglichkeiten

- **Alphabetische Liste**
  - *Alle Begriffe* des Wörterbuchs einschließlich der "freien" Schlagwörter sollen sich als alphabetische Liste anzeigen lassen.
  - *"Freie" Schlagwörter* sollen gekennzeichnet sein.
  - *Belegungshäufigkeit:* Zu jedem Begriff soll die Belegungshäufigkeit (Anzahl der Literaturnachweise, die diesen Begriff als Schlagwort enthalten) angegeben werden.

  Es wäre hilfreich, wenn das Modul in der Lage wäre, alle Begriffe, die einer bestimmten *Trunkierung* entsprechen, in alphabetischer Folge anzuzeigen. Für rechtstrunkierte Begriffe wird diesem Wunsch durch die normale alphabetische Anzeige entsprochen, die Anzeige sollte aber auch bei Links- bzw. Links-und-Rechts-Trunkierung möglich sein.

- **Strukturanzeige**. Es soll möglich sein, zu jedem Begriff - durch direkte Eingabe des Begriffs oder durch Markieren in der alphabetischen Liste bzw. in der Strukturanzeige - sich dessen Umfeld zeigen zu lassen. Zu diesem *Umfeld* gehören:
  - eine eventuelle textliche *Erläuterung*
  - *Oberbegriff(e)* (der nächsthöheren Ebene)
  - *Unterbegriffe* (der nächsttieferen Ebene)
  - *Synonyme*
  - *assoziierte Begriffe*

- Begriffe, die durch *weitere Relationen* [s.o. zu Gruppenbildung] zugeordnet sind.

Außerdem sollte die *Belegungshäufigkeit* angezeigt werden.

Aus der Anzeige sollte erkenntlich sein, ob ein Begriff am *Anfang oder Ende einer Hierarchieleiter* steht; anders gesagt: Es sollte zu erkennen sein, ob ein angezeigter Oberbegriff noch Oberbegriffe über sich oder ein angezeigter Unterbegriff noch Unterbegriffe unter sich hat. Dieselbe Strukturanzeige soll *auch beim Aufruf eines Synonyms* erscheinen.

- **Navigieren**. Es soll möglich sein, sich die Begriffe der höchsten Ebene anzeigen zu lassen und von da aus (durch Markieren/Anklicken eines Begriffs) sich auf einzelnen Zweigen der Hierarchie nach unten und erforderlichenfalls auch wieder nach oben oder *über assoziierte Begriffe in andere Hierarchiezweige* zu bewegen.

  Es soll möglich sein, jederzeit und *unter Beibehaltung des aktuell gewählten Begriffs zwischen alphabetischer Anzeige, Strukturanzeige und Navigieren* zu wechseln.
- **Kopieren.** Es soll möglich sein, Begriffe in der Anzeige zu markieren, die dann automatisch *in die Recherchefrage* übernommen werden.

1.3 Recherchemöglichkeiten

- **Synonyme**. Neben der Recherche nach der *Äquivalenzklasse insgesamt* soll es auch möglich sein, nur nach einem *einzelnen Element*, einem bestimmten Synonym zu recherchieren.
- **Automatische Erweiterung nach unten**. Es soll möglich sein, wahlweise alle *Unterbegriffe* (bis zum Ende der Hierarchieleiter) automatisch in die Recherche einzubeziehen.

  Es soll möglich sein, die mittels der freien Relation [s. o. zu 1.1 "Gruppenbildung"] hierarchieübergreifend zu einer *Gruppe* zusammengefaßten Begriffe wahlweise automatisch in die Recherche einzubeziehen.
- **Trunkierung (Maskierung)**. Es soll *zwei Joker* geben: einen für ein Zeichen, einen für mehrere (oder kein) Zeichen.

  Trunkierung soll uneingeschränkt *rechtsseitig, linksseitig sowie beidseitig* und auch bei Zeichenketten, die Leerzeichen enthalten, zulässig sein. Der Joker für ein Zeichen soll auch *innerhalb* der Zeichenkette benutzt werden dürfen.

  Trunkierung soll sowohl bei der *direkten* als auch bei der *sequentiellen* Recherche möglich sein.

2 Diskussion einiger Probleme

2.1 Vorbemerkung

Im folgenden sind einige Aspekte der aufgeführten Anforderungen herausgegriffen mit dem Ziel, die Problematik anschaulicher werden zu lassen; eine tiefergehende Diskussion läßt der vorgegebene Rahmen nicht zu. Natürlich stehen die genannten Anforderungen nicht nur untereinander in Zusammenhang, sondern haben auch zu anderen Elementen der Recherchesoftware Beziehung, zu den Möglichkeiten der Volltextrecherche beispielsweise, und darüber hinaus zu grundlegenden Faktoren, wie die ins Auge gefaßte Nutzerklientel und andere Intentionen des Datenbasenherstellers. Hierauf wird aber nicht eingegangen.

2.2 Synonyme

Bei der heutigen Leistungsfähigkeit auch einfacher Computersysteme kann - bei Vorliegen eines strukturierten Wörterbuchs - die Zusammenführung bedeutungsähnlicher Benennungen ausschließlich dem Computer überlassen werden. Damit ist die früher übliche und auch heute noch oftmals praktizierte Differenzierung in erlaubte Deskriptoren und verbotene Nichtdeskriptoren überholt. Dies führt zu zwei wesentlichen Vorteilen: Das Dokument kann "konkret" indexiert werden, das heißt, zur Kennzeichnung des Inhalts darf ein sehr spezifischer Begriff benutzt werden, und - was in Anbetracht der zunehmenden Endnutzernähe solcher Systeme vielleicht von noch größerer Bedeutung ist - das System kann so eingerichtet werden, daß jeder Benutzer für seinen Einstieg in das System das Wortgut vorfindet, das ihm aus seinem Berufs- und Erfahrungskreis am ehesten geläufig ist. Letzteres spielt gerade bei Datenbanken eine Rolle, die für Benutzer sehr unterschiedlicher Berufe und Bereiche gedacht sind. Die projektierte Literaturdatenbank Berufliche Bildung ist eine solche Datenbank.

Für die Recherchesoftware bedeutet dies:

- Automatische Zusammenführung von Begriffen zu Äquivalenzklassen und Möglichkeit, nach der Äquivalenzklasse insgesamt zu recherchieren.
- Möglichkeit, auch nach jedem Synonym einzeln zu recherchieren, um eine Untermenge spezifischerer Dokumente erhalten zu können (Zusammenhang mit der Methode "konkreter" Indexierung).
- Transparenz der Äquivalenzklassen, das heißt, der Benutzer muß nachsehen können, welche Begriffe zur Äquivalenzklasse gehören (siehe auch 2.6).

2.3 Freie Schlagwörter

Bei Datenbanken, die durch das Zusammenführen von Nachweisen aus unterschiedlichen anderen Datenbanken aufgebaut werden, wie dies bei der geplanten Literaturdatenbank Berufliche Bildung der Fall sein wird, ist bei jeder Lieferung mit dem Auftauchen neuer Schlagwörter zu rechnen. Ein Wörterbuch, das unter solchen Bedingungen benutzt wird, muß besonders lebendig und flexibel sein. So darf es keine Probleme bereiten, neue Begriffe in das Wörterbuch aufzunehmen oder die Struktur zu verändern. Für die Recherche- und Verwaltungssoftware bedeutet das:

- Automatische Aufnahme neuer Begriffe in das Wörterbuch;
- Anzeige dieser Begriffe im alphabetischen Wörterbuch mit einer Kennung, die besagt, daß der Begriff (noch) nicht in die Wörterbuchstruktur eingebaut ist;
- Möglichkeit des Ausdrucks einer alphabetischen Liste dieser Begriffe als eine Voraussetzung für die Wörterbuchpflege (für die Vorbereitung der Einordnung in die Struktur).

2.4 Assoziationen

Jeder Benutzer bewegt sich in relativ gefestigten Begriffs- und Denkmustern, die u. U. von den Mustern des Indexierers oder des Datenbank-Verantwortlichen deutlich abweichen. Assoziationen helfen, von einem Muster ins andere zu finden. Assoziationen werden vielfach auch aufgrund von Erfahrungen der Indexierer oder professionellen Rechercheure eingebaut. Sie nehmen dem Benutzer einen Teil des Nachdenkens (wo er denn wohl noch suchen könnte) ab bzw. vermeiden Suchen am falschen Platz. Weil assoziative Verweise deutliche Hilfen sind, sich in einem strukturierten Wörterbuch zu bewegen und Rechercheverluste zu vermeiden, soll die Software assoziative Beziehungen verwalten. Eine wahlweise automatische Einbeziehung assoziativer Begriffe in die Recherche erscheint aber nicht erforderlich, da diese Siehe-auch-Verweise immer der individuellen Auswahl bedürfen.

2.5 Freie Relationen

Eine zu weitgehende hierarchische und insbesondere polyhierarchische Gliederung eines Wörterbuchs kann dazu führen, daß der Sinn der Gliederung ins Gegenteil verkehrt wird: Das Wörterbuch wird unübersichtlich. Auch aus diesem Grunde sind freie Relationen nützlich, die eine Gliederung von Wörterbucheintragungen außerhalb der primären Hierarchiestruktur ermöglichen. Die Anzeige solcher Strukturen weist den recherchierenden Benutzer auf be-

stimmte Zusammenhänge hin; lassen sich die Strukturen für die Recherche aktivieren, so erspart das dem Rechercheur die Eintragung der Einzelbegriffe. *Bild 1* zeigt, wie dem Begriff ALTE BUNDESLÄNDER die einzelnen Bundesländer - von BAYERN bis SCHLESWIG-HOLSTEIN - zugeordnet werden, ohne eine zusätzliche Hierarchiestufe einzufügen. Noch deutlicher wird das Prinzip in *Bild 2* am Beispiel der Gruppen EU-LÄNDER und OECD-LÄNDER, wobei die Gruppe OECD-LÄNDER ja noch auf weitere, im Bild nicht dargestellte Hierarchiezweige, wie z. B. ASIEN oder AMERIKA, zugreift.

2.6 Länge des Wörterbuchbegriffs

Oftmals ist die Länge der Wörterbucheintragung begrenzt, beispielsweise auf 60 oder gar nur 40 Zeichen. Zwar hat die überwiegende Mehrzahl der Wörter noch deutlich weniger Zeichen, muß man aber ein längeres Wort kürzen, so ist dies um so ärgerlicher. Am häufigsten davon betroffen sind Namen von Institutionen. Vor allem der gelegentliche Benutzer hat Mühe, einen gesuchten Begriff, der im Wörterbuch gekürzt ist, zu finden, manchmal auch bei der Entschlüsselung einer Kürzung. Es gibt heute keinen überzeugenden Grund mehr, die Länge einer Eintragung so stark einzuschränken.

2.7 Strukturanzeige

Für den Benutzer wird der Thesaurus interessant, wenn er sich in ihm bewegen kann und aus ihm Anregungen zur Verbesserung seiner Fragestellung erhält. Die Anzeige des Wörterbuchs und seiner Strukturen hat einen doppelten Sinn: Zum einen soll sie helfen, relevante, dem Suchziel am besten dienende Suchbegriffe zu finden, zum anderen soll sie den Thesaurus und über ihn zugleich die Arbeit des Recherchesystems transparent machen. Wenn zum Beispiel das Recherchesystem nach einer Äquivalenzgruppe recherchiert, so muß dem Recherchierenden die Möglichkeit gegeben sein, die Elemente dieser Gruppe zu erfahren; er darf dem System nicht blindlings ausgeliefert sein.

Viele überflüssige Eingaben oder Eingabeversuche könnten dem Recherchierenden erspart werden, wenn aus der Strukturanzeige immer zu erkennen wäre, ob ein angezeigter Unterbegriff weitere Begriffe unter sich hat.

3 Einige Beispiele

Im folgenden werden einige Beipiele aus verfügbaren PC-Datenbanken bzw. entsprechender Software gebracht, die eher zufällig sind. Sie haben hier keinen anderen Sinn, als die Berechtigung der vorgetragenen Forderungen und Wünsche beispielhaft zu veranschaulichen und zu unterstreichen. Mit diesen

Beispielen ist keine kritische Auseinandersetzung mit der erwähnten Software verbunden, denn eine solche Bewertung müßte auch andere Aspekte berücksichtigen.

Die CD-ROM-Datenbank Literaturdokumentation Bildung[2] benutzt keinen Thesaurus, intern werden aber Synonyme verwaltet. Die Synonymverwaltung arbeitet verdeckt und erlaubt keinen Einblick in die Zusammensetzung der Synonymgruppe. So ist der Benutzer auf Probieren angewiesen, was Zeit kostet und unter Umständen auch frustriert. Ein Beispiel zeigt *Bild 3*. An dieser

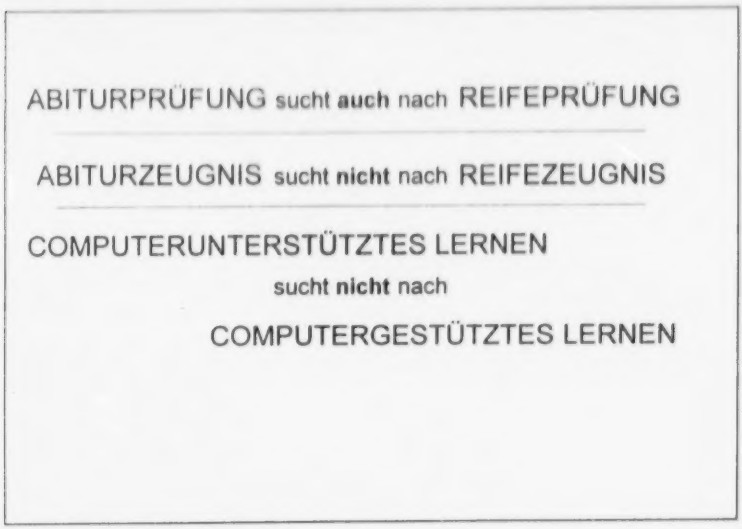


Bild 3 Beispiele der Zusammenfassung von Begriffen bei verdeckt geführten Synonymgruppen

Stelle geht es nicht um die Frage, ob die Zusammenfassung oder Nicht-Zusammenfassung bestimmter Begriffe zu einer Äquivalenzklasse dem einen eher sinnvoll erscheint und dem anderen weniger, sondern darum, daß dem Recherchierenden die Kontrolle darüber entzogen wird, nach welchen Begriffen er tatsächlich recherchiert und welche ihm wichtig erscheinenden Begriffe möglicherweise nicht in die Suche einbezogen sind. Das Beispiel zeigt, wie

[2] Herausgeber: FIS Bildung; Software: CD Answer, V3.01, von Dataware Technologies, Inc.

notwendig eine Offenlegung der Synonymbeziehungen wäre. Eine Suche nach dem einzelnen Synonym ist nicht möglich. Suchbegriffe können aus der alphabetischen Liste in die Suchfrage kopiert werden.

Die Software ODARS, die u. a. für CD-ROM **REHADAT** des Instituts der deutschen Wirtschaft verwendet wird, hat für das Feld "Suchbegriffe" ebenfalls eine Synonymverwaltung. Anders als bei CD Answer kann man hier die Synonymgruppe einsehen (*Bild 4* und *5*). Die Häufigkeitsangabe beim Suchwort bezieht sich auf das Synonym, nicht auf die gesamte Äquivalenzklasse. Eine Recherche nach dem einzelnen Synonym ist aber dennoch nicht möglich.

Obwohl die Datenbank **KURS DIREKT**[3] Weiterbildungsveranstaltungen ausweist, ist ihre Struktur mit der einer Literaturdatenbank vergleichbar. In dieser Datenbank gibt es eine Systematik; sie kann als Thesaurus interpretiert werden. Jedem verbalen Begriff ist gleichberechtigt eine Notation zugeordnet; andere Synonyme gibt es nicht. Die Begriffe sind hierarchisch in einer Baumstruktur gegliedert. Zu zahlreichen Begriffen gibt es assoziative Verweise. Beim Start einer Recherche werden in jedem Fall automatisch alle zu einem Begriff gehörenden Unterbegriffe bis zum Ende der Hierarchieleiter mit eingeschlossen. Die in den Abbildungen (*Bild 6* bis *8*) zu erkennenden Tasten ("Tiefer", "Höher") gestatten das Navigieren, auch hin zu Assoziationen (Taste "Verweis"). Vor dem Suchwort steht in jedem Fall die Belegungshäufigkeit. Die Taste "Übernahme" kopiert den Begriff in die Rechercheфrage.

Die vom Landesinstitut für Schule und Weiterbildung in Soest[4] auf CD-ROM herausgegebene **Literaturdatenbank Bildungswesen** benutzt als Software CICADE-Multimedia (Version 3.01 für Windows). Über eine Schaltfläche kann der DOPAED-Thesaurus eingesehen werden, und es lassen sich Begriffe markieren und in die Rechercheфrage kopieren. Für die Auswahl des Wortguts ist es vorteilhaft, im Thesaurus blättern und in den Strukturen frei navigieren zu können. Der Thesaurus läuft aber völlig autonom *neben* dem Recherchesystem; dadurch wird der Benutzer auch zu fehlerhaften Schlüssen und Eingaben verleitet. So sind im Thesaurus Synonyme ausgewiesen, und in der Thesaurusanzeige steht beispielsweise (*Bild 9* bis *11*):

REIFEPRÜFUNG - benutzt für **ABITUR**

[3] Herausgeber: Bundesanstalt für Arbeit, Nürnberg, und Institut der deutschen Wirtschaft, Köln; Software KOMPASS für CD

[4] Landesinstitut für Schule und Weiterbildung, Referat Z3, Paradieser Weg 64, 59494 Soest

1HILFE 2 3SYNONY 4 5 6 7 8 9 10

LITERATUR

	Suchbegriffe	
Eingabedatum	LEHRABBRUCH	1
Referenz-Nr.	LEHRABSCHLUß	3
	LEHRAMT	1
Suchbegriffe	LEHRE	5
	LEHRER	16
	LEHRGANG	17
	LEHRGANGSBEWERTUNG	1
Dokumentart	LEHRGANGSLEITER	1
Autor	LEHRGANGSPROGRAMM	1
Herausgeber	LEHRLING	2
Titel	LEHRMETHODE	2
Übertitel	LEHRMITTEL	1
Zeitschrift	LEHRPLAN	3
Jahr,Heft	LEHRSTUHL	1
Jahr	LEIBSCHMERZ	1
Verlag,Ort	LEICHTBAU	1
Text	LEICHTBAUWEISE	1
	LEIDEN	2
	LEIDENSMINDERUNG	1

REHADAT CD-ROM SUCHERGEBNIS:

Bild 4 Alphabetische Anzeige der Suchbegriffe mit Belegungshäufigkeit bei ODARS

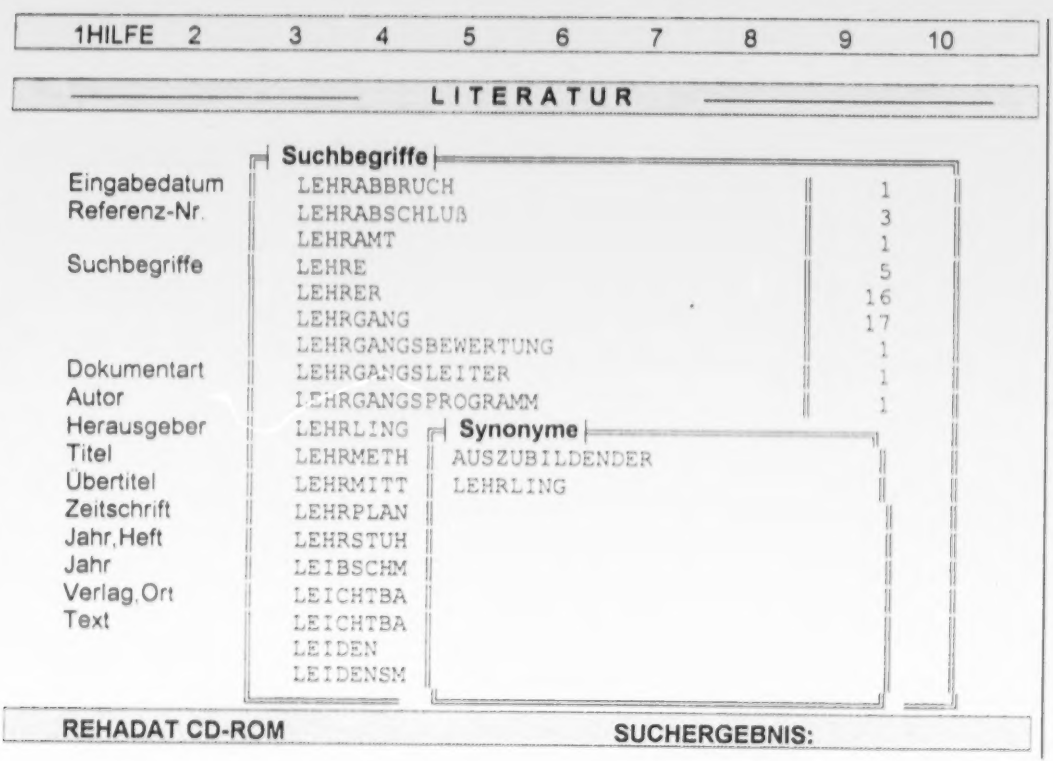


Bild 5 Anzeige aller Synonyme zum Stichwort LEHRLING

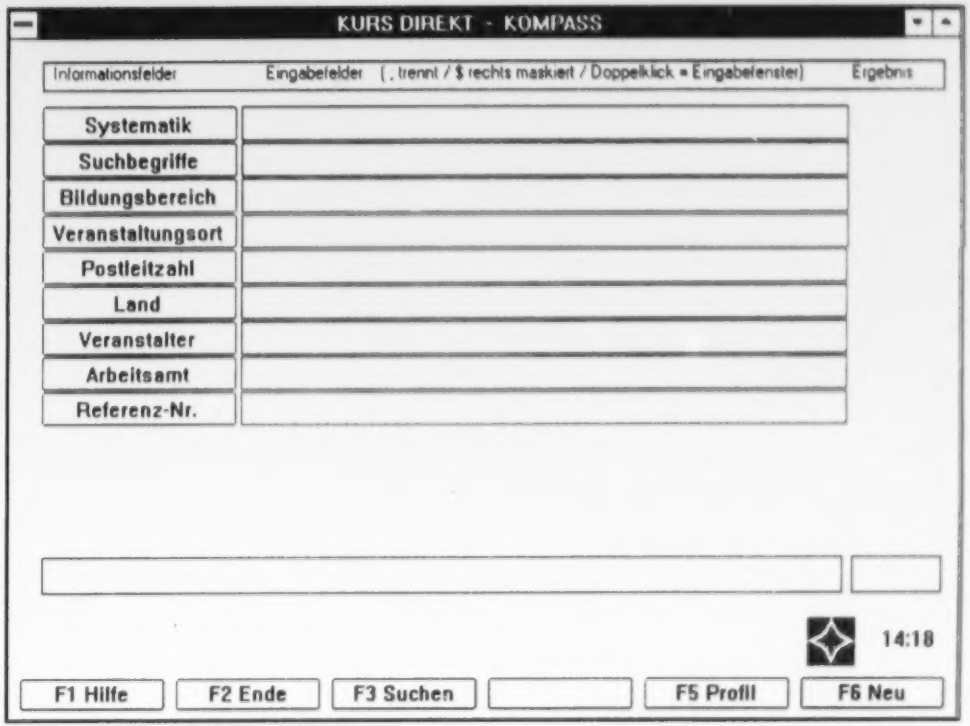


Bild 6 *KURS DIREKT. Eroffnungsbildschirm*

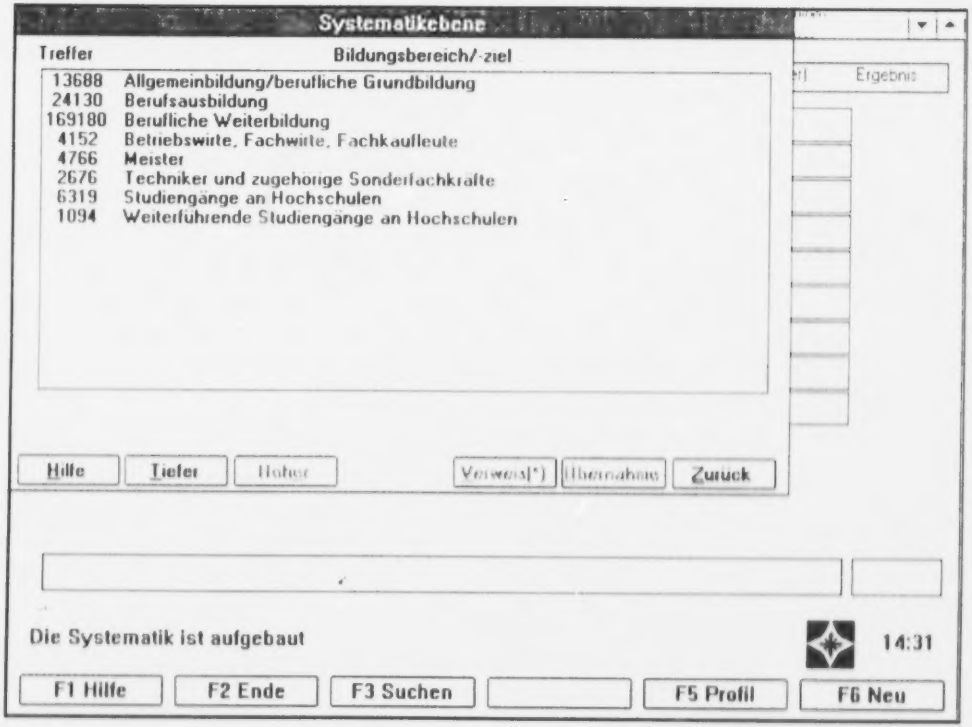


Bild 7 Anzeige der ersten Ebene

Bild 8 Bewegen im "Thesaurus"

BEST COPY AVAILABLE

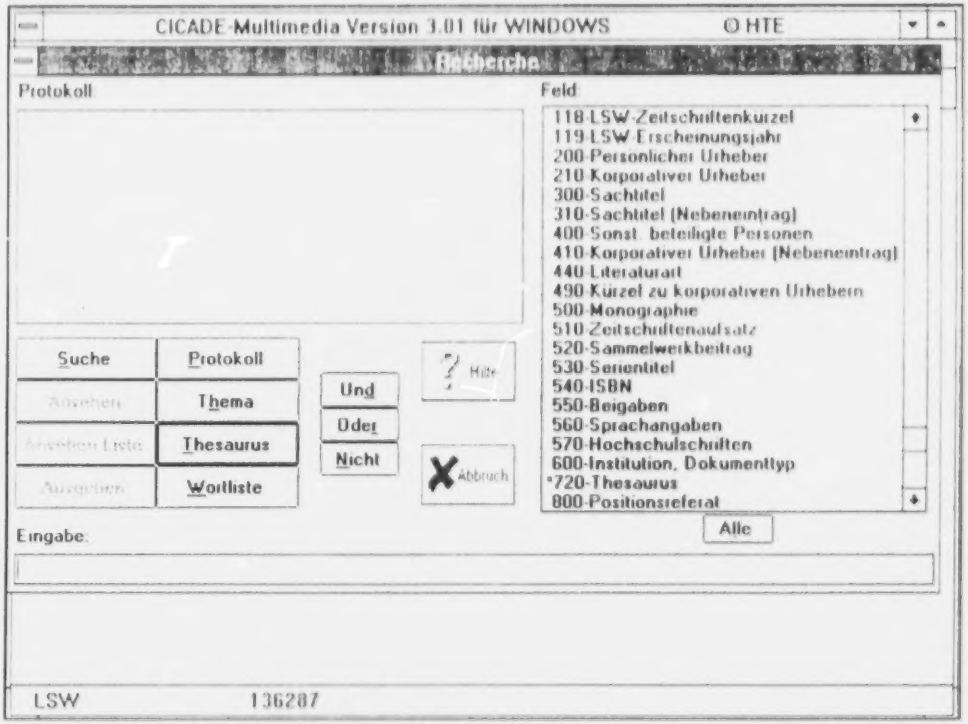


Bild 9 Eröffnungsbildschirm. Feld 720-Thesaurus als Suchfeld aktiviert

Bild 10 Strukturanzeige zum Begriff REIFEPRÜFUNG - hiernach steht REIFEPRÜFUNG auch für ABITUR

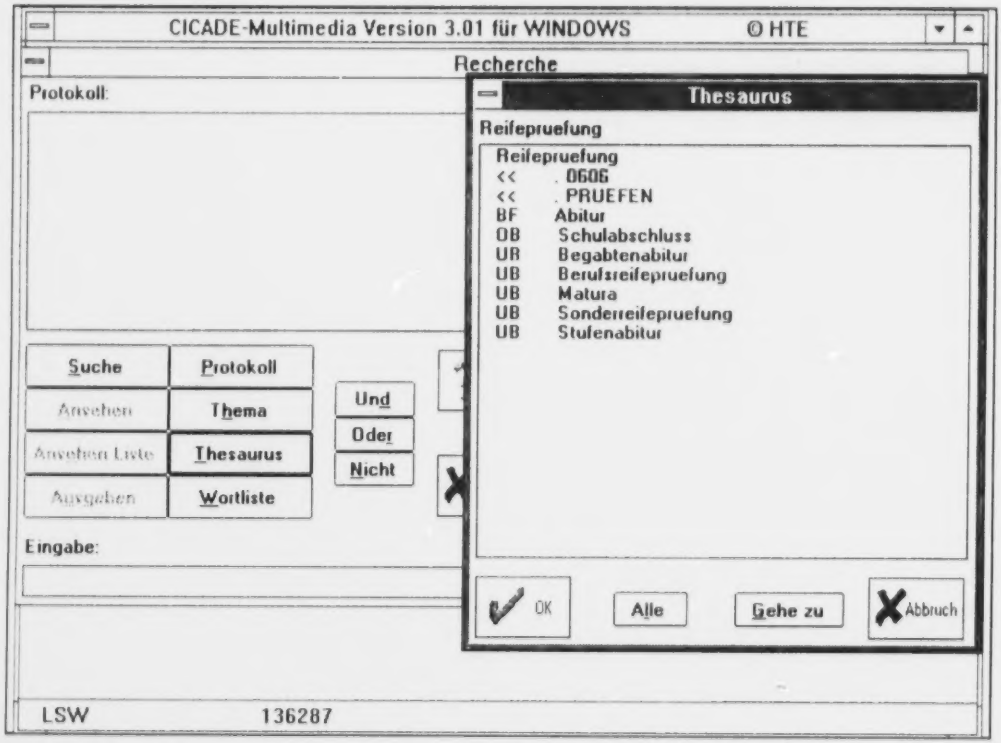


Bild 11 Rechercheergebnisse - sie zeigen:
Die in Bild 10 angekündigte Verknüpfung wird nicht realisiert.

Das ist aber eine rein theoretische Fiktion. In der Datenbank selbst werden die beiden Begriffe völlig gleichartig und isoliert voneinander behandelt. Wie das Rechercheergebnis ausweist, sind sie nicht miteinander verknüpft: Die Suche nach REIFEPRÜFUNG liefert 371 Nachweise, nach ABITUR weitere 17 Nachweise (ein Nachweis erhält beide Suchwörter im sog. "Thesaurus"-Feld). Verläßt man sich also auf die Aussagen des Thesaurus, so führt das unter Umständen zu beträchtlichen Rechercheverlusten. Dieses Beispiel unterstreicht besonders anschaulich die Notwendigkeit, den Thesaurus und seine Strukturen tatsächlich in die Recherchesoftware zu integrieren.

Das verbreitete System **LARS**[5] verwaltet in der Standardversion keinen Thesaurus, jedoch wird in der speziellen Bibliotheksanwendung ein Thesaurusmodul angeboten. Die folgenden, zum Vergleich genannten Angaben, beruhen auf Informationen des Herstellers. Danach wird die Hierarchie durch Notationen, die der Anwender zusätzlich vergeben muß, organisiert. Polyhierarchie ist möglich. Neben der Hierarchie wird auch Synonymie und Assoziation verwaltet. Alle diese Beziehungen können eingesehen, Unterbegriffe und Assoziationen wahlweise in die Recherche automatisch einbezogen werden. Neben dem Thesaurus wird eine Indexliste aufgebaut. Im Thesaurusfeld vorhandene, aber noch nicht in die Thesaurusstruktur eingefügte Begriffe erscheinen nur dort. Da sie nicht besonders gekennzeichnet sind, ist die Thesauruspflege möglicherweise erschwert. Auch die Belegungshäufigkeit wird nur in der Indexliste (die also nicht identisch ist mit der alphabetischen Thesaurusliste) angezeigt, nicht im Thesaurus. Vorausgesetzt, der Thesaurus ist tatsächlich integriert - und nicht nur, wie bei der CICADE-Version - danebengestellt, würde das beschriebene LARS-Modul dennoch einen wesentlichen Teil der Anforderungen erfüllen.

Für das ebenfalls verbreitet angewendete **LIDOS**[6] der Doris Land Software-Entwicklung, Oberasbach, war gerade die Thesaurusfunktion ein Element, das LIDOS vor anderen auszeichnete. Deshalb wurde auch LIDOS (Vers. 3.3) hinsichtlich einer Verwendung für die Literaturdatenbank Berufliche Bildung geprüft. Von den Eigenschaften seien mit Bezug auf das behandelte Thema die folgenden genannt: Die Länge der Thesaurusbegriffe ist auf nur 40 Zeichen begrenzt, die Belegungshäufigkeit ist aus der Anzeige nicht zu ersehen, hierarchieübergreifende assoziative Beziehungen werden nicht verwaltet, freie Schlagwörter im Thesaurusfeld verhindern das Laden des entsprechenden Nachweises. Dies alles sind Eigenschaften, die, wie eingangs dargelegt wurde, aus unserer Sicht unbefriedigend sind und in einem positiven Sinne verändert werden sollten.

Abschließend soll kurz das Programmsystem vorgestellt werden, das uns bei der Entwicklung der Literaturdatenbank Berufliche Bildung unterstützt. Das

[5] Herausgeber: Lars GmbH, Frankfurt am Main
[6] Herausgeber: Doris Land Software-Entwicklung

AIDOS/M (Version 3.1)[7] der DataPrint GmbH, Dresden, ist hinsichtlich der Wörterbuchverwaltung und des Instrumentariums für die Eingabeprüfung von Nachweisen das mächtigste uns bekannte PC-Programm für Literaturverwaltung und Recherche dieser Preisklasse (etwa 3.200,- DM). Wir benutzen das Programm für den Aufbau von Testdatenbanken und Thesaurus. Das Programm verwaltet in einer Datenbank 10 autonome Datenbasen (mit je max. 64.000 Nachweisen), jede mit maximal 31 strukturierten Wörterbüchern und 31 Klassifikationen. Im Wörterbuch sind Oberbegriffe, Unterbegriffe, Synonyme, Assoziationen und 10 weitere, frei definierbare symmetrische oder gerichtete Relationen und auch textliche Anmerkungen möglich. Ein Wörterbuchbegriff darf 60 Zeichen lang sein. Das Wörterbuch läßt sich alphabetisch und systematisch - jeder Begriff mit seinem Umfeld - anzeigen; dabei wird immer zu jedem am Bildschirm sichtbaren Begriff die Belegungshäufigkeit mit angezeigt. Es besteht die Möglichkeit, den Zugang für Nachweise mit freien Schlagwörtern zu sperren oder zu öffnen. Läßt man freie Schlagwörter zu, so werden diese in der alphabetischen Liste besonders gekennzeichnet, und sie lassen sich auch gesondert ausdrucken, was die Thesauruspflege erleichtert. Bei der Recherche, die gleichzeitig über mehrere Datenbasen laufen kann, lassen sich alle Thesaurusbeziehungen automatisch aktivieren. Trunkierung ist rechts und links auch mehrfach möglich. Hinsichtlich Thesaurusverwaltung und Recherche genügt dieses DOS-Programm nahezu allen eingangs genannten Anforderungen. Würde das Programm von einigen Mängeln in anderen Bereichen befreit und in einer Windows-Version vorliegen, wäre es sicherlich eine für unsere Zwecke beinahe ideale Software.

Anschrift des Verfassers:

Dr. Heinz Ziegler
Straße der Pariser Kommune 36
10243 Berlin
Tel. (030) 294 73 77

[7] Herausgeber: DataPrint GmbH Dresden

Yrla M. Labouvie / Petra Konz

Qualitätskriterien und Qualitätssicherungsmaßnahmen, aufgezeigt an der Datenbank PSYNDEX

1 Qualitätskriterien für die inhaltliche Abbildung von Literaturdokumenten in einer Datenbank

Ohne die einzelnen Qualitätskriterien ausführlich diskutieren zu wollen, werden nachfolgend Qualitätskriterien für die inhaltliche Abbildung von Literaturdokumenten genannt.

1.1 Multiplizität der inhaltlichen Erschließung

Die inhaltliche Abbildung sollte in einer Vielzahl von Kategorien erfolgen, die sich sinnvoll ergänzen. Dies erhöht die Chance, den Inhalt einer Publikation vollständig und treffend zu beschreiben.

1.2 Inhaltliche Zuverlässigkeit

- Terminologische Kontrolle
  Besonders in der Psychologie ist es erforderlich, den Wortschatz für die Beschreibung der Literaturdokumente zu kontrollieren, indem die Fachbegriffe eindeutig definiert sind, ihr Umfang begrenzt, aber dennoch so umfassend ist, daß das Fachgebiet treffend, vollständig und aktuell repräsentiert wird.
- Richtigkeit
  Der Inhalt des Dokumentes ist korrekt darzustellen.
- Genauigkeit
  Der Inhalt des Dokumentes ist treffend darzustellen.
- Vollständigkeit
  Alle relevanten Inhalte des Dokumentes sind aufzuführen.

- Konsistenz
  Die Fachterminologie muß insbesondere bei der Indexierung sowohl interals auch intraindividuell gleichbedeutend verwendet werden. Dies beinhaltet Einheitlichkeit und Kontinuität bei der Verwendung der Terminologie.
- Objektivität
  Die relevanten Inhalte sind ohne Wertung darzustellen.
- Verständlichkeit
  Die Inhalte müssen so beschrieben werden, daß sie auch ohne spezifische Fachkenntnisse verstehbar sind.

1.3 Formale Zuverlässigkeit

- Konsistenz
  Die Schreibweise der Fachbegriffe muß festgelegt sein; die Darstellung der Inhalte sollte konsequent den festgelegten Standards und Strukturen folgen; darüber hinaus sollte die Differenziertheit der inhaltlichen Erschließung für alle Dokumente gleich sein.
- Korrektheit
  Rechtschreibfehler sollten ausgemerzt werden.

1.4 Mehrsprachigkeit

- Terminologische Kontrolle
  Die Übersetzung der Fachterminologie, gleich in welche Sprache, muß soweit wie möglich kontrolliert sein, d. h. inhaltlich treffend und korrekt.
- Formale Konsistenz
  Die Schreibweise der Übersetzung muß klar festgelegt sein; darüber hinaus sollte Kontinuität bei der Mehrsprachigkeit einzelner Felder gewahrt bleiben.
- Formale Korrektheit
  Rechtschreibfehler in der Übersetzung müssen vermieden werden.

2. Maßnahmen der Qualitätssicherung

2.1 Definition von Beschreibungskategorien

Für die Referenzdatenbank PSYNDEX stehen für die inhaltliche Abbildung der Literaturdokumente die folgenden Kategorien zur Verfügung:

- Thesaurus-Deskriptor,
- "Freier" Deskriptor,
- "Phrase" (= kurze Annotation),
- Klassifikation,
- Kurzreferat.

Diese Kategorien ergänzen einander; sie entsprechen damit dem Kriterium "Multiplizität der inhaltlichen Erschließung". Vier dieser Kategorien sind Pflichtfelder, womit "Kontinuität in der Differenziertheit der Erschließung" erreicht wird.

2.2 Instrumente und Richtlinien der inhaltlichen Erschließung

Für die einzelnen Kategorien sind Instrumente und Richtlinien entwickelt worden, die die inhaltliche Zuverlässigkeit, formale Konsistenz und Mehrsprachigkeit unterstützen, wenn nicht gar garantieren.

Im folgenden werden zu den einzelnen Kategorien spezielle Maßnahmen der Qualitätssicherung aufgeführt.

Definition von Thesaurus-Deskriptoren

Die Dokumente der Datenbank PSYNDEX werden mit dem amerikanischen "Thesaurus of Psychological Index Terms", herausgegeben von der American Psychological Association, verschlagwortet. Dieser Thesaurus, der in seiner Terminologie die gesamte Psychologie abzudecken versucht (Kriterium: Vollständigkeit der Terminologie), umfaßt in seiner siebten Version 7.311 Begriffe: rund 4.300 aktive Deskriptoren, knapp 200 inaktiv gewordene Deskriptoren und 2.800 Synonyme bzw. Quasi-Synonyme. Er zeigt im Hauptteil (Relationship section) die Begriffe in ihrer hierarchischen Struktur, in einem Netzwerk von Ober- und Unterbegriffen und verwandten Begriffen. Zu einer Vielzahl von Deskriptoren werden sogenannte "scope notes" aufgeführt (Kriterium: Eindeutigkeit der Begriffe). Darüber hinaus nennt der Thesaurus sämtliche Begriffe nach ihrem Wortstamm alphabetisch rotiert (Rotated alphabetical terms section); außerdem bietet er zu ausgewählten Themen Cluster inhaltlich zusammengehörender Deskriptoren (Cluster section); zum Beispiel gibt es einen "Educational Cluster".

Von der Zentralstelle für Psychologische Information und Dokumentation (ZPID) wurde das gesamte Thesaurus-Vokabular ins Deutsche übersetzt (Kriterium: Mehrsprachigkeit), darüber hinaus wurden deutsche Synonymbe-

griffe festgelegt, die auf den deutschen Deskriptor verweisen (Kriterium: Terminologische Kontrolle der Übersetzung).

Der Thesaurus wird alle drei bis vier Jahre aktualisiert.

Eindeutigkeit, Vollständigkeit und Aktualität des kontrollierten Vokabulars unterstützen die Qualitätskriterien: Richtigkeit, Genauigkeit, Vollständigkeit und Konsistenz der inhaltlichen Erschließung. Die Übersetzung der Terminologie sichert Mehrsprachigkeit. Die Festlegung der Schreibweise der Begriffe unterstützt die formale Konsistenz.

Definition von "Freien" Deskriptoren

Die "Freien" Deskriptoren repräsentieren Konzepte, die im Thesaurus nicht enthalten sind (Kriterium: Vollständigkeit des kontrollierten Vokabulars). So erfaßt der Thesaurus etwa aufgrund unterschiedlicher Traditionen und Forschungsschwerpunkte der deutschsprachigen gegenüber der amerikanischen Psychologie spezifische deutsche Aspekte der Psychologie - insbesondere das Bildungssystem und Testverfahren - nur unzureichend. Ferner kann die alle drei bis vier Jahre erfolgende Revision des Thesaurus mit der wissenschaftlichen Entwicklung nicht ausreichend Schritt halten, so daß neue Inhalte zumindest anhand vorläufiger Deskriptoren abgebildet werden müssen. Die Definition "Freier" Deskriptoren fördert die Genauigkeit, Aktualität und Vollständigkeit der inhaltlichen Erschließung.

Bei den "Freien" Deskriptoren handelt es sich um Schlagwörter ohne Anbindung an die hierarchische Struktur des Thesaurus. Häufig sind sie nur zeitlich begrenzt eingeführt, eben nur so lange, bis dieser Begriff als Thesaurus-Deskriptor eingeführt wird. Insgesamt stehen für PSYNDEX etwa 200 "Freie" Deskriptoren zur Indexierung zur Verfügung; fast alle sind ins Englische übersetzt (Kriterium: Mehrsprachigkeit).

Indexierungsregeln

Um eine hohe Indexierqualität zu erreichen, sind bei der Vergabe von Thesaurus-Deskriptoren verschiedene Regeln zu beachten. Die wichtigsten sind:

- Es muß mindestens ein Thesaurus-Deskriptor (Pflichtfeld, garantiert Kontinuität in der Feldbelegung), es können maximal 17 zugeordnet werden.
- Es soll eine möglichst spezifische Indexierung mit Vergabe der präzisesten Begriffe erfolgen (Kriterium: Genauigkeit).
- Die dem Dokument zugeordneten Thesaurus-Deskriptoren sind zu differenzieren in (standardisierte Darstellung):

  - Hauptdeskriptoren (maximal sechs), sie stehen für die Inhalte mit zentraler Bedeutung;
  - Nebendeskriptoren (maximal elf), sie stehen für zusätzlich in der Publikation angesprochene Aspekte.
- Die Thesaurus-Deskriptoren sollen inhaltlich abbilden:
  - alle Begriffe bzw. Konzepte im Titel,
  - unabhängige/abhängige Variablen bei empirischen und experimentellen Untersuchungen; bei sonstigen Publikationen die zentralen, insbesondere die psychologischen Konzepte,
  - Methoden und Verfahren,
  - die untersuchten Populationen,
  - Personennamen,
  - geographische Namen,
  - Testnamen,
  - Medikamentennamen,
  - den formalen Charakter der Publikation.

Diese Vorgaben stützen die Vollständigkeit der inhaltlichen Erschließung.

- "Freie" Deskriptoren sind nur ergänzend zu den Thesaurus-Deskriptoren zu verwenden, nur dann, wenn der entsprechende Sachverhalt in der Publikation von besonderer Bedeutung ist.

Regeln für die Formulierung der "Phrase"

Die "Phrase" stellt eine kurze Beschreibung (Annotation) des Inhaltes der Publikation dar. Dieser unkontrollierte Wortschatz soll die Deskriptoren präzisieren und ergänzen (Kriterium: Genauigkeit und Vollständigkeit). Die "Phrase" führt Inhalte, die durch Deskriptoren nicht oder nicht ausreichend beschrieben werden können, in strukturierter, die Zusammenhänge verdeutlichender Form auf. Eine Wiederholung von Deskriptoren ist nur erwünscht, wenn diese bereits einen hohen Präzisionsgrad aufweisen. Die "Phrase" dient mehr dem Auffinden eines Literaturnachweises, im Gegensatz etwa zum Abstract, das primär auf die Informationsrepräsentation des Dokumentes abzielt.

Die "Phrase" wird ausschließlich in englischer Sprache formuliert.

In einem Auswertermanual wird versucht, soweit wie möglich Vorgaben zur Strukturierung und Standardisierung der "Phrase" zu machen. Sie unterstützen die formale und inhaltliche Konsistenz.

> Beispiel: age-correlated limits of cognitive development capacity, intervention study on acquisition of memory task, testing-the-limits approach, 35 young & aged adults, experimental study.

Definition einer Sachgebietsklassifikation

Der Datenbank PSYNDEX liegt das Klassifikationssystem der Datenbank PsycINFO zugrunde. Es handelt sich um ein dreistufiges monohierarchisches Klassifikationssystem, bestehend aus insgesamt 157 Klassifikationskategorien mit 22 Haupt- und 135 Untergruppen. Zur Definition der einzelnen Kategorien stehen ausführliche "scope notes" (Kriterium: Terminologische Kontrolle) zur Verfügung. Im Jahr 1991 ist das Klassifikationssystem grundlegend überarbeitet und verbessert worden (Kriterium: Aktualität).

Von der ZPID wurde das Klassifikationssystem ins Deutsche übersetzt (Kriterium: Mehrsprachigkeit).

Die englische und deutsche Schreibweise der Klassifikationsgruppen ist eindeutig festgelegt (Kriterium: Terminologische Kontrolle auf formaler Ebene).

Klassifizierungsregeln

Bei der Einordnung eines Literaturdokumentes in das Klassifikationssystem sind folgende Gesichtspunkte zu beachten:

- Das Dokument ist mindestens in eine, höchstens in vier Klassifikationsgruppen einzuordnen.
- Die ausgewählten Klassifikationskategorien sind zu differenzieren in:
  - die relevanteste Kategorie (entspricht dem Hauptschwerpunkt der Publikation),
  - nachrangige Kategorie/n (stehen für den untergeordneten Schwerpunkt der Publikation).
- Das Klassifizieren soll so spezifisch und präzise wie möglich erfolgen, also sind in erster Linie die Untergruppen der Klassifikation zu verwenden.

Regeln zur Formulierung des Kurzreferates

- Abstractart (Vorgabe fördert formale Konsistenz).
  Das Abstrahieren von Zeitschriftenartikeln, Dissertationen, Reports und Sammelwerksbeiträgen erfolgt in informativen Referaten (ohne Bewertung). Nach DIN 1426 soll ein entsprechendes Abstract "insbesondere Auskunft über das behandelte Gebiet, Zielsetzungen, Hypothesen, Methodologie, Ergebnisse und Schlußfolgerungen ... geben". D. h. es werden beim informativen Referat auch konkrete Resultate (quantitive und qualitative Da-

ten) aufgeführt. Bücher (Monographien, Sammelwerke) werden in indikativen Referaten dargestellt, in denen auf die im Originaldokument behandelten Sachverhalte aufzählend hingewiesen wird.
- Das Abstract soll kurz, prägnant, in einem wissenschaftlich-fachlichen Stil formuliert sein. Redundanz ist zu vermeiden, ebenso wie journalistische Redewendungen.
- Für die verschiedensten Publikationsarten, wie empirische Studien, experimentelle Studien, methodische Studien, theoretische Studien, Übersichtsarbeiten, Fallberichte, Erfahrungsberichte, Bibliographien, Test- und Verfahrensentwicklungen, Biographische Studien, Nachrufe, Kritiken und Repliken, Diskussionen, Interviews, populärwissenschaftliche Beiträge, liegen spezielle Gliederungsregeln vor. Diese Vorgaben fördern formale Konsistenz, Vollständigkeit und Genauigkeit der inhaltlichen Beschreibung.
- Bei der Abstrahierung gelten gewisse "stilistische Konventionen". Sie legen fest: Zeitformen, passive Redewendungen, Aufzählungen, Nennung von Zahlen, Test-, Fragebogen- und Verfahrensnamen, Medikamentenbezeichnungen, Angaben von Korrelation und Signifikanzniveau, Abkürzungen sowie Sonderzeichen (Kriterium: Formale Konsistenz/standardisierte Darstellungsform).

Automatische Prüfläufe und Ergänzungen bei der Erfassung und den Updates

Beim Update werden die Pflichtfelder auf Belegung geprüft. Dies garantiert Vollständigkeit der Dokumentbeschreibung und Kontinuität der Feldbelegung.

Das gesamte Thesaurus-Vokabular (deutsch + englisch) ist in einer Hilfsdatei an den Erfassungspool angebunden; diese Hilfsdatei schließt Rechtschreibfehler bei der Eingabe der englischen Terms absolut aus. Wird ein Deskriptor falsch (Rechtschreibfehler) oder nicht in der definierten Schreibweise eingegeben, wird der Eintrag vom System abgelehnt. Die deutschen Begriffe werden beim Update aus der Hilfsdatei automatisch zugespielt. All dies garantiert formale Korrektheit der Terms sowie ihre standardisierte Schreibweise.

Die "Freien" Deskriptoren, die bei der Erfassung sowohl deutsch als auch englisch eingetippt werden müssen, sind in keiner Hilfsdatei hinterlegt, so daß eine automatische Vergleichsprüfung mit der Schlagwortliste nicht erfolgen kann. Ein individuelles Korrekturlesen ist hier erforderlich, was keine absolute Zuverlässigkeit garantiert. Eine Standardisierung der Schreibweise über eine Hilfsdatei ist geplant.

Das gesamte Klassifikationssystem (englisch + deutsch), einschließlich der Klassifikationscodes, ist als Hilfsdatei an den Erfassungspool angebunden.

Die Eingabe der Klassifikationsgruppe in die Erfassungsmaske kann als vierstelliger numerischer Code oder im englischen Wortlaut (Trunkierung möglich) erfolgen. Ein nicht vorhandener Code wird bei der Eingabe nicht akzeptiert. Über den Code in der Hilfsdatei wird der Dokumentbeschreibung der volle Wortlaut der Klassifikationsgruppe deutsch und englisch zugespielt, damit sind Rechtschreibfehler ausgeschlossen. Hilfsdatei und Prüfprogramme sichern einheitliche Schreibweise und Übersetzung sowie formale Korrektheit.

Übergreifende Maßnahmen zur Qualitätssicherung

Eine Reihe übergreifender personeller und organisatorischer Maßnahmen kommen der Qualität der inhaltlichen Erschließung zugute.

Die "Inhaltliche Zuverlässigkeit" stützen vor allem folgende Maßnahmen:

- hohe fachliche Qualifikation der Dokumentare,
- "native speakers" für Übersetzungen sowie Auswertungen in englischer Sprache,
- soweit möglich Berücksichtigung der Spezialgebiete der Auswerter,
- Autopsieprinzip, d. h. die Publikation muß im Original für die Auswertung vorliegen,
- Erfahrungen der Auswerter mit Datenbankrecherchen,
- schneller Zugriff des Auswerters auf PSYNDEX während der inhaltlichen Erschließung,
- namentliche Verantwortung des Auswerters für seine inhaltliche Erschließung,
- Schulung der Auswerter.

Übergreifende Maßnahmen, die die "formale Zuverlässigkeit" stützen, sind mehrfache abteilungsinterne Korrekturläufe.

3 Datenbank-Struktur vs. Datenbank-Design

In diesem Kapitel soll aufgezeigt werden, inwieweit die vom Produzenten eingebrachte Datenbankstruktur und -qualität der inhaltlichen Erschließung von verschiedenen Hosts in den Datenbankdesigns umgesetzt worden sind.

3.1 GRIPS-Version der Datenbank PSYNDEX (Host DIMDI)

- Bei der Umsetzung ins GRIPS-Format ist die vom Produzenten vorgegebene analytische Struktur der Datenerfassung erhalten geblieben. Jeder Kategorie entspricht ein Datenbankfeld. Der Wortschatz ist in jedem Feld absuchbar, anzeigbar und ausdruckbar.

Kategorie	*GRIPS-Feld*
Thesaurus-Deskriptor englisch	Controlled term (CT)
Thesaurus-Deskriptor deutsch	Controlled term German (CTG)
"Freier" Deskriptor englisch	Index term (IT)
"Freier" Deskriptor deutsch	Index term German (ITG)
Klassifikationsgruppe englisch	Section heading (SH)
Klassifikationsgruppe deutsch	Section heading German (SHG)
Phrase	Key phrase (KP)
Kurzreferat englisch	Abstract English (ABE)
Kurzreferat deutsch	Abstract German (ABG)

- Die Kategorien sind auf Feldebene unterschiedlich invertiert. CT, CTG, IT, ITG, SH, SHG sind sowohl direkt- als auch freitext invertiert, andere ausschließlich im Freitext ansteuerbar: /AB, /KP.
- Ausschließlich die inhaltlichen Felder sind im Basic Index abgelegt, d. h. der Freitext umfaßt: TIG, TIE, ABG, ABE, CT, CTG, SH, SHG, IT, ITG, KP. Also auch SH und SHG; dies kann bei der Suche zu irrelevanten Nachweisen führen. Beispielsweise wenn man im Freitext nach "movement therapy" sucht, qualifizieren sich gleichermaßen Nachweise zu "art therapy" über die Klassifikationsgruppe "Art & Music & Movement Therapy". Dieser Mangel wird bei der Reduktion der Freitextsuche über den "Preferred Qualifier" (/PQ) wiederum ausgeglichen.
- Die Mehrsprachigkeit der Erschließung ist voll und differenziert erhalten geblieben; dies bietet die Möglichkeit, das GRIPS-Kommando sprachspezifisch auszurichten und zwar mit Hilfe von Sprach-Qualifiers.

  /ENGL umfaßt: ABE, CT, IT, TIE, KP;
  /GERM umfaßt: ABG, CTG, ITG, TIG.

Diese Differenzierung auf Sprachebene ist insbesondere deshalb so wichtig, um Suchstrategien zwischen PSYNDEX und PsycINFO übertragen zu können. Auch im Rahmen der Superbase-Funktion sollte die Suchsprache englisch formuliert sein.

- Der Thesaurus ist im Display in seiner vollen hierarchischen Struktur online anzeigbar (mit "history" und "scope notes" und Übersetzungen), differen-

ziert nach deutschem und englischem Wortschatz, ebenso wie der rotierte alphabetische Teil des Thesaurus.
- Das Find-Kommando bietet darüber hinaus die Möglichkeit, die hierarchische Struktur in die Deskriptor-Suche einzuschließen (Down-, Up-, Related-, Tree-Funktion). Zusätzlich wird die Deskriptor-Suche insofern unterstützt, als bei Eingabe eines englischen oder deutschen Synonymbegriffes automatisch der (Haupt-)Deskriptor abgesucht wird.
- Das Klassifikationssystem ist in seiner hierarchischen Struktur gleichfalls deutsch oder englisch online anzeigbar. Im Find-Kommando werden auch hier über die Down-, Up- bzw. Tree-Funktion Unter- bzw. Hauptgruppen automatisch in die Suche eingeschlossen.
  Darüber hinaus sind die Klassifikationscodes integriert; sie ermöglichen eine Anzeige der Klassifikation in ihrer hierarchischen Struktur.
- Die bei der Erschließung vorgenommene Differenzierung in Haupt- und Nebendeskriptoren sowie in Haupt- und Nebenklassifikationsgruppe ist im Datenbank-Design erhalten geblieben. Mit Hilfe des Gewichtungs-Qualifier läßt sich dieser Aspekt in die Suche einbringen:

  Weighting /w1 anwendbar bei: CT, CTG, SH, SHG

3.2 SPIRS-Version der Datenbank PSYNDEX (SilverPlatter CD-ROM)

Die SPIRS-Version bietet einen der GRIPS-Version ähnlich differenzierten Zugang auf PSYNDEX, d. h. die analytische Kategorienstruktur, auch auf Sprachebene, ist erhalten geblieben. Der Basic Index umfaßt in der PC-SPIRS-Version den Wortschatz aus allen inhaltlichen, bibliographischen und formalen Feldern.

Der Thesaurus ist ausschließlich in englischer Version auf der CD-ROM anzeigbar; explosion-ähnliche Funktionen sind möglich. Das Klassifikationssystem ist bei der WinSPIRS-Version (1.0) anzeigbar und wird dies auch bei der PC-SPIRS-Version (3.3) sein. Die Gewichtung der Thesaurus-Deskriptoren sowie der Klassifikationsgruppen ist in die Suche nicht einschließbar.

3.3 AOS-Version der Datenbank PSYNDEX (Host GBI)

In der AOS-Version (vormals SUDOK) ist die Mehrsprachigkeit der Erschließung zwar erhalten geblieben, aber keinesfalls in der differenzierten Form wie bei GRIPS und SPIRS. Hier sind die Kategorien Thesaurus-Deskriptor deutsch bzw. englisch in einem Feld zusammengefaßt, ebenso die Kategorien Klassifikationsgruppe deutsch bzw. englisch. Sprachqualifier sind

nicht einsetzbar. Die Kategorien "Freier" Deskriptor deutsch bzw. englisch sind nicht implementiert.

Der Thesaurus sowie das Klassifikationssystem sind in ihren Strukturen nicht integriert, Explosion-Funktionen daher nicht möglich. Gewichtete Suchen nach Thesaurus-Deskriptoren bzw. Klassifikationsgruppen sind ausgeschlossen.

Ein Display auf Feld-Ebene ist nur auf Einzelwort-Ebene möglich; diese Reduktion schränkt den gewünschten Überblick über Thesaurus-Deskriptoren erheblich ein.

Der Basic Index umfaßt wie die PC SPIRS-Version den Wortschatz sowohl der inhaltlichen als auch der bibliographischen und formalen Felder.

Abschließend kann festgestellt werden, daß die GRIPS-Version der Datenbank PSYNDEX die vom Produzenten angestrebte Qualität der inhaltlichen Abbildung der nachgewiesenen Literaturdokumente nahezu vollständig im Datenbank-Design umsetzt. Die PC-SPIRS-Version und vor allem die Win-SPIRS-Version der PSYNDEX-CD-ROM kommen der GRIPS-Version am nächsten. Hingegen bleibt das Datenbank-Design der AOS-Version hinter den Möglichkeiten, die die inhaltliche Erschließung der Literaturdokumente bietet, am weitesten zurück.

Literatur

Bonfig, B. (1994): PSYNDEX Indexing & Abstracting Manual (unveröffentlichtes Manuskript). Trier: Zentralstelle für Psychologische Information und Dokumentation, Universität Trier.

Bonfig, B. (1993): ZPID Style & Terminology Guidelines for PSYNDEX (unveröffentlichtes Manuskript). Trier: Zentralstelle für Psychologische Information und Dokumentation, Universität Trier.

Cazan, C. (1993): Qualitätskontrolle großer bibliographischer Datenbanken am Beispiel EMBASE und BIOSIS nebst vier Statements zur aktuellen Diskussion. In: Wolfram Neubauer (Hrsg.): Qualität und Information. Deutscher Dokumentartag 1993. Frankfurt: Deutsche Gesellschaft für Dokumentation, S. 223 - 230.

Genth, P. (1993): Qualitätsmanagement in einem Host-Betrieb. In: Wolfram Neubauer (Hrsg.): Qualität und Information. Deutscher Dokumentartag 1993. Frankfurt: Deutsche Gesellschaft für Dokumentation, S. 49 - 55.

Henzler, R. (1992): Information und Dokumentation. Berlin: Springer.

Hobohm, H-C. (1993): Online versus print: Medienspezifische Qualitätsanforderungen im Zeitalter des Mehrwerts. In: Wolfram Neubauer (Hrsg.): Qualität und Information. Deutscher Doku-

mentartag 1993. Frankfurt: Deutsche Gesellschaft für Dokumentation, S. 123 - 134.

Käfer, G. (1993): Qualitätsstrategien der Anbieter: Qualitätsüberlegungen bei Juris. *Password*, 9, S. 17-18.

Klassen-Meurer, J. (1993): Qualitätskontrolle am Beispiel der FAKT-Datenbank. In: Wolfram Neubauer (Hrsg.): Qualität und Information. Deutscher Dokumentartag 1993. (S. 507-517). Frankfurt: Deutsche Gesellschaft für Dokumentation.

Marbach, J. (1992): Erfahrungen mit CD-ROM. Bewertungskriterien in Theorie und Praxis. In: Wolfram Neubauer/Susanne Plagemann (Hrsg.): 14. Online-Tagung der DGD. Frankfurt am Main, 27. bis 30. April 1992. Proceedings. Frankfurt: Deutsche Gesellschaft für Dokumentation, S. 27 - 38.

Müller, J. (1977): Leitfaden für die inhaltliche Erschließung von Zeitschriftenaufsätzen. Trier: Zentralstelle für Psychologische Information und Dokumentation, Universität Trier (z. Zt. in Überarbeitung).

Reinert, G. & Wolff, U. (1978): Richtlinien und Hinweise für das Verfassen von Kurzreferaten über psychologische Dissertationen. Trier: Zentralstelle für Psychologische Information und Dokumentation, Universität Trier.

Ruffer, J. (1989): Qualitätskriterien innerbetrieblicher Informationsvermittlung. *Nachrichten für Dokumentation*, 5, S. 309-312.

Schwuchow, W. (1993): Qualitätsmanagement für Informationsdienste. Philosophie der 90er Jahre? *Cogito*, 1, S. 2-7.

Stroetmann, K. (1993): Qualität von Informationsdiensten als strategische Managementaufgabe. In: Werner Schwuchow (Hrsg.): Qualität von Informationsdiensten. Frankfurt: Deutsche Gesellschaft für Dokumentation, S. 10 - 22.

Walker Jr., A. (Hrsg.) (1994): Thesaurus of psychological index terms (7. Aufl.). Washington: American Psychological Association.

Zentralstelle für Psychologische Information und Dokumentation. (1994): PSYNDEX Terms: Deskriptoren / Subject Terms zu den Datenbanken PSYNDEX und PSYTKOM (4., aktual. Aufl.). Trier: Zentralstelle für Psychologische Information und Dokumentation, Universität Trier.

Anschrift der Verfasserinnen

Yrla M. Labouvie/Petra Konz
ZPID Universität Trier
Universitätsring 15
54286 Trier
Tel. (06 51) 201-2877; Fax: -2071

Walther Umstätter

Bewertung der Qualität von Datenbanken

Benutzeroberfläche

Liest man einen Titel, wie den obenstehenden, so hinterfragt man natürlich zunächst, was verstehen wir in diesem Zusammenhang unter einer **Datenbank**? Martha Williams, die bekannte Professorin aus den USA, die auch Herausgeberin eines bekannten Database Directories ist, hat einmal unterschieden zwischen "Databases" - entsprechend den Literaturdatenbanken - und "Databanks" - entsprechend den Faktendatenbanken. Die Assoziation der verwandten Base hat in Deutschland vermutlich die Benutzung des Wortes Datenbase etwas gebremst. Datenbanken enthalten bekanntlich Dokumente, sogenannte Records, die durch eine Retrievalsoftware gezielt auffindbar sind. Zu den Literatur- und Faktendatenbanken gesellen sich von Jahr zu Jahr mehr Volltextdatenbanken, bei denen es schwierig wird zwischen elektronischen oder digitalen Büchern bzw. Zeitschriften einerseits und Datenbanken andererseits zu unterscheiden. Auch wenn diese Feinheiten hier für uns keine Rolle spielen, so sind sie juristisch nicht unwichtig. Das digitale Buch als eigenständige Publikation war bislang genau betrachtet eine Volltextdatenbank mit zunehmend multimedialem Charakter.

Daneben gewinnen aber **Wissensbanken** mehr und mehr an Bedeutung, bei denen hier für uns insbesondere der Bereich der sogenannten Intelligent Interfaces von Interesse ist. Diese Wissensbanken enthalten z. B. kausal oder zumindest heuristisch verknüpfte Informationen darüber, wie man bestimmte Datenbanken am besten nutzen kann. Damit haben wir als Informationsvermittler die Möglichkeit, unsere Erfahrung auf eine Benutzeroberfläche zu transferieren, die es auch dem Neuling ermöglicht, dieses unser Wissen zu nutzen. Insbesondere beim Einsatz von CD-ROMs, wo der Zeitdruck und die Übertragungsmenge eine viel geringere Rolle spielt als in den klassischen Online Retrieval Systemen über Datenfernzugriff, können Guided Tours durch die Informationslandschaft sehr hilfreich sein.

Hinsichtlich der **Qualitätskontrolle**, die in der Informations- und Dokumentationslandschaft insofern eine besondere Rolle spielt, als der Anteil an soge-

nanntem Informationsmüll besonders groß ist, müssen wir zwei Ebenen unterscheiden:

1. Die Qualität der **Datenbankinhalte**; so machen z. B. Schreibfehler, überflüssige Wiederholungen, Irreführungen, unzureichende Erschließung, mangelhafte Vollständigkeit etc. dem Rechercheur das Leben schwer.
2. Die Qualität von **Datenbankprogrammen**; hierzu gehören die Benutzeroberfläche, das Retrieval und die Leistungsfähigkeit der Input/Output-Funktionen, um nur drei zu nennen.

Schreibfehler in der Größenordnung von einem Fehler auf 2000 Zeichen sind durchaus normal. Bei besonders schlampigen Angeboten finden wir allerdings im Freitext auch dann nichts, wenn statt "Schulbildung" in der Datenbank "Schul- bildung " zu finden ist, weil im Orginaltext an dieser Stelle ein Zeilenwechsel vorlag. Daß in amerikanischen Datenbanken deutsche Worte und Namen besonders häufig falsch geschrieben sind, kann nicht verwundern. Trotzdem kann man sich über Fehler in der Größenordnung von weniger als einem Prozent nicht erregen, wenn man aus langjähriger Erfahrung weiß, daß eine durchschnittliche Recherche in einer der klassischen Datenbanken wie ERIC, MEDLARS, BIOSIS etc. nur grob geschätzt 50% Precision und 50% Completeness bzw. Recall Ratio erbringt.[1] Trotzdem hat man bei OCLC, dem größten Online-Katalogverbund der Welt, schon vor Jahren Software zum Einsatz gebracht, die für eine bestimmte Qualitätskontrolle sorgt, weil Verbundkataloge unter einer besonders hohen Fehlerrate leiden. Dies liegt in erster Linie an den recht komplexen Katalogisierungsregeln, wie wir sie bei AACR2 und RAK kennen.

Mit anderen Worten, herkömmliche Retrievalsysteme mit Bool'scher Logik und indexierten Dokumenten aus Zeitschriften, korrekt als DBE (Dokumentarische Bezugseinheiten) bezeichnet, erbringen etwa zur Hälfte unbrauchbare Dokumente und etwa die Hälfte dessen, was wir an Literatur zu einem Thema finden müßten. Bei OPACs (Online Public Access Catalog) liegt der Wert sogar noch weit niedriger.

Hier setzt natürlich die Qualität der **Erschließung** ein, die seit den Erfahrungen der Cranfield Studie in den frühen 60er Jahren, auch durch geschickte Thesauri oder Facettenklassifikationen nicht signifikant verbessert werden kann.[2]

Im Gegensatz zur landläufigen Ansicht, daß eine sachliche Erschließung mit Hilfe eines guten Thesaurus mehr Information bringt, zeigten mir vor mehre-

[1] W. Umstätter: Nutzen der Indexierung bei Online-Datenbanken. In: 14. Online Tagung der DGD Proceedings, Frankfurt/M. DGD-Schrift (OLBG-13) 2 (1992), S. 403-420

[2] F. W. Lancaster: Information Retrieval Systems, 2. Aufl., New York (Wiley) (1979)

ren Jahren eigene Untersuchungen deutlich, daß ein guter Thesaurus vielmehr zu einem gleichmäßigen bzw. mittelmäßigen Ergebnis führt. Dies ist auch verständlich, da Thesaurusbegriffe die Indexierung insofern vereinheitlichen, als sie bestimmte Spezialbegriffe nicht zulassen und andere zu allgemein und häufig benutzte Begriffe einschränkend spezifizieren.

Beim klassischen Online-Retrieval mit hohen Recherchekosten war und ist ein solches System, das zuverlässig mittelmäßige Rechercheergebnisse bringt, höchst vorteilhaft. Es verhinderte in den Zeiten, als die meisten Endnutzer noch keine Ahnung vom Online-Retrieval hatten, daß wir für teures Geld zu dem Ergebnis kamen, dem Endnutzer keine relevanten Treffer geben zu können. Sobald uns der Thesaurus anzeigte, daß bestimmte Begriffe in ihm verzeichnet waren, konnten wir auch mit einer entsprechenden Wahrscheinlichkeit ein brauchbares Rechercheergebnis erwarten. Für die Vollständigkeit einer Suche ist diese Erschließungsstrategie von Nachteil.

Lediglich Volltexte erlauben eine vergleichsweise hohe Recall Ratio, wenn man über ausreichend leistungsfähige Retrievalsysteme verfügt und natürlich auch über die notwendigen fachlichen Kenntnisse, um in ihnen recherchieren zu können. Eine häufig zu beobachtende Schwierigkeit, der auch erfahrene Onliner unterliegen, besteht in der Verwendung von typischen Suchbegriffen und Deskriptoren, wie wir sie bei indexierten Systemen einzusetzen gewohnt sind. Sie führen bei Volltextrecherchen natürlich nicht zum gewohnten Erfolg. Dagegen sind Volltexte, und Retrievalsysteme mit intelligenter Benutzeroberfläche gut geeignet, dem Endnutzer beim Recherchieren zu helfen, und den weitaus größeren Gehalt an Information mittels Textanalysen zu erschließen. Das ist kein Vergleich z. B. zu gängigen OPACs mit durchschnittlich 1,7 Schlagwörtern pro Buch.

Die sogenannten **Intelligent Interfaces** sind dazu angetan, die Rechercheergebnisse zumindest durch eine gute Benutzerführung zu verbessern. Dies kann beispielsweise durch ein Relevance Feedback geschehen, obwohl seit langem bekannt ist, daß die Kenntnisse des erfahrenen Onliners beim Retrieval unverzichtbar sind. Sein Wissen in sogenannte Knowledge Workstations zu implementieren ist daher die neue Strategie des Information Resources Managements.

Hier sind zunächst auf der untersten Stufe die **Kontextbezogenen Hilfefunktionen** zu nennen, die z. B. bei Windows-Programmen meist mit F1 zu erreichen sind und deren Standardisierung sehr zu begrüßen ist. Trotzdem birgt jede Standardisierung früher oder später einen Mangel an Innovation in sich und muß nach einer gewissen Zeit aufgegeben werden. Außerdem kann ein Wettbewerb der Qualität nur außerhalb des Standards stattfinden. So hat sich bekanntermaßen Apple mit seiner Lisa- bzw. Mac-Philosophie lange vom Einsatz der Funktionstasten distanziert und konsequent die Benutzung der Maus bevorzugt.

Der Ruf nach Standardisierung ist alt und brachte zunächst in Europa die sogenannte Common Command Language (CCL) hervor, eine vereinheitlichte Retrievalsprache, die sich jedoch über DIMDI, ESA, ECHO und einige andere Hosts hinaus nicht sehr weit verbreiten konnte. Der Versuch der Hosts, sich gerade über eine hervorstechend bessere Retrievalsprache zu profilieren, verhinderte die Akzeptanz der CCL. Erst durch INTERNET kam man in den USA zu der Erkenntnis, im Rahmen der ANSI Norm Z39.50 eine CCL einzuführen, weil der Endnutzer oft nicht weiß, in welchem Rechner er sich gerade befindet und mit welchen Befehlen er arbeiten kann. Diese neue CCL (NISO Z39.58) hat große Ähnlichkeit mit dem was wir hier in Deutschland von DIMDI kennen, ist aber keinesfalls identisch. Sie wird auch von vielen INTERNET-Rechnern nur fragmentarisch erfüllt. Sie kennt Befehle wie:

BAC Back	CHO Chose	DEF Define	DEL Delete
DIS Display	EXP Explain	FIN Find	HEL Help
MOR More	PRI Print	REL Relate	REV Review
SAV Save	SCA Scan	SET Set	SHO Show
SOR Sort	STA Start	STO Stop	

Tab. 1: Befehle der CCL nach NISO Z39.58

Bei einem rasanten Fortschritt der Datenbanktechnologie und -software, wie wir ihn heute beobachten, kann unsere Hoffnung unmöglich in einer Standardisierung liegen, auch wenn sie uns zwischenzeitlich hilft, mit den Systemen umzugehen.

Benutzeroberflächen unterliegen meist dem **Test durch bestimmte Zielgruppen**. Diese Zielgruppen haben eine vergleichbare Erfahrung auf ihrem Fachgebiet, im Umgang mit Rechnern und eine mehr oder minder ähnliche Einstellung beim Einsatz von Rechentechnik, die sich allerdings fast überall und oft unbemerkt systemimmanent ändert. So konnte man beim Auftauchen des ersten Apple Lisa-Rechners noch erhebliche Ressentiments gegenüber dem Einsatz von Icons, Rollbalken und Maus beobachten. Inzwischen haben wir uns fast alle an das Drag-and-Drop gewöhnt, wie es nun auch Windows anbietet.

Daß Icons der geschriebenen Information gegenüber weit überlegen sind, wenn man ihre Bedeutung kennt und mit ihnen täglich umgeht, kann der Anfänger natürlich auch noch nicht abschätzen.

Von fundamentaler Bedeutung ist für jede Hilfestellung und jede Benutzerführung die **stringente Philosophie**, die hinter einem System steht. Sobald diese Philosophie vom Endnutzer einmal erkannt und verstanden worden ist, hat er meist keine Probleme, ihr konsequent zu folgen. Insofern haben sich auch Oberflächen weit überlebt, die einmal Funktionstasten, dann wieder be-

stimmte Tastenkombinationen bis hin zur sogenannten Geierkralle und zwischendurch, ganz unerwartet, die Maus erfordern.

Umgekehrt ist natürlich der optimale Einsatz verschiedener Eingabemöglichkeiten bei unterschiedlichen Erfordernissen inzwischen ebenso Usus wie die Verwendung von Farbe, Bild und Ton. Hier kommt es allerdings immer wieder zu grundsätzlichen Systemwechseln, die oft automatisch zu radikalen Veränderungen der Systemphilosophie führen.

Verfolgte man früher meist die Philosophie, menügesteuert möglichst viel Information dem Endnutzer zur Auswahl anzubieten, so ist heute zunehmend der Trend zu beobachten, das System selbst nach der optimalen Lösung gegebener Probleme suchen zu lassen. Damit haben wir einen zwingenden Übergang von den sogenannten dummen zu den intelligenten Oberflächen, die auf ihr gespeichertes Wissen zurückgreifen, d. h. auf kausal und heuristisch verknüpfte Information.

Hier greift insbesondere der **informationswissenschaftliche Ansatz**, bei dem ein Maximum an Information auf den Bildschirm gebracht werden muß. Dies darf allerdings nicht mit einem Maximum an Zeichen bzw. dem sogenannten **Information Overload** verwechselt werden, das eher dem sogenannten Noise entspricht. So kann ein überladener Bildschirm, bei dem der Endnutzer nur mit großem Aufwand die Information herausfindet, die er braucht, um an dieser Stelle weiter zu arbeiten, eher mit dem informationstheoretischen Rauschen einer Telefonleitung verglichen werden als mit einer klar erkennbaren und verständlichen Nachricht. Der gezielte Blick auf einen bestimmten Punkt, von dem man weiß, daß er die gewünschte Information enthält, der rasch und

Beenden | Hilfe | Inhaltsverzeichnis | Nachschlagen | Neue Datenbank Rückwärts | Suche | Thesaurus | Vorwärts | Wähle | Zeigen
?
Ausgabe | Auswahl | Bildschirmausgabe | Datei | Druck | Einstellungen Erklärung | Kommunikation | Löschen | Nachbarschaftssuche Parameter setzen | Schrifttyp | Sortieren | Speichern | Starten | Stop

Abb. 1: Typische Bildschirmoberfläche alter Prägung mit alphabetischen Wörterlisten, die unübersichtlich und oft wenig aussagekräftig waren.

klar erkennbar ist, und der sofort signalisiert, was hier erlaubt und verboten ist, kann nur erfolgen, wenn wir aus Erfahrung wissen, wo wir was zu erwarten haben.

Auch die Vergrößerung solcher Bildschirme bringt wenig. So kennen wir beispielsweise eine tote Vergrößerung, die trotz Größenzunahme keine zusätzliche Information bringt. Dies gilt sowohl für eine gut lesbare Schrift, die durch Vergrößerung nicht mehr aussagt, als auch für grafische Darstellungen, bei denen die Vergrößerung nur zu einer Art Würfelgrafik führt. Dagegen sind sogenannte Pull Down Menus als inhaltliche Vertiefungen bzw. Vergrößerungen anzusehen, die bekanntermaßen recht hilfreich sein können.

Alphabetische Listen von Wörtern, die, aus dem Zusammenhang gerissen, kaum zu verstehen sind, helfen nur wenig weiter. Die Nutzung der Informationssyntaktik, bei der jedes neu hinzukommende Bit die doppelte Menge der Information enthält, die vorher empfangen worden ist, kann teilweise genutzt werden, indem syntaktische Entscheidungsbäume die Oberfläche transparent machen. Je mehr Rückgriffe auf vorhandenes Wissen beim Endnutzer getätigt werden können, desto mehr Information ist über den Bildschirm oder den Lautsprecher vermittelbar. Dabei kann eine eingängige Farb- und Positionssymbolik von erheblichem Vorteil sein. Beide sollten allerdings ebenso einer einheitlichen Philosophie folgen.

Einschlägige Versuche zum Problem der Icons bzw. Pictogramme und ihres Einsatzes sind mehrfach in der Welt gemacht worden. So auch bei dem System Bookhouse, im Rahmen eines ESPRIT-Projektes zur Benutzung einer virtuellen Bibliothek. Hier zeigte sich deutlich, daß eine verbale Erklärung der Icons zumindest am Anfang unverzichtbar ist.[3]

Intelligente Systeme sind auch in der Lage, Menüketten wie Trampelpfade durch einen Informationswald nutzerspezifisch umzugestalten. Dies erfreut sich insbesondere bei INTERNET im sogenannten Gopherspace einer breiten Beliebtheit.

Sowohl die Datenbanken selbst als auch die intelligenten Oberflächen gehen heute in die Richtung der natürlichen Sprache. Die Eingabe einer einfachen Frage, wie: "The role of women in World War II" und ihre direkte Beantwortung durch Literaturzitate mit ähnlichen Titeln, durch ein System wie den Knowledge Finder,[4] sind dafür typisch.

---

[3] MOHAWK - Modelling of Human Activities in Work Context' CEC ESPRIT Basic Research Action 3105, Taxonomy for Analysis of Work Domains. MOHAWK Workshop Liege 15. - 16. 05. 1990 Vol. 1 RISOE National Laboratory Roskilde, Denmark (1990)

[4] Knowledge Finder ist ein Programm der Firma Aries Systems, USA, das in Deutschland von Nova Idea in Köln verbreitet wird.

Als **Kriterien für bedienungsfreundliche Systeme** lassen sich heute die folgenden aufzählen:

- klar verständliche und spezifisch relevante *Fehlermeldungen*, die auch den Ausweg aufzeigen,
- hypertextartige Hilfefunktionen, die das Handbuch in verständlicher Sprache einschließen,
- einfach zu erstellende *Makros*, die stumpfsinnige Routinen verkürzen und die im Rahmen einer Scriptsprache auch nachträglich editierbar sind,
- Tastenkombinationen, sogenannte *Short cuts*, die den direkten Aufruf von Befehlen erleichtern,
- ein annehmbares *Zeitverhalten*,
- gleichzeitiges Angebot von *Menüführung* und *Expertenmodus* für Anfänger und Routiniers,
- Einsatz von *Fenstertechnik*, die es jederzeit erlaubt, aus einer vorgegebenen Situation auszubrechen,
- Angebot von grafischen *Elementen*, die eine Standortbestimmung erleichtern,
- Berücksichtigung der *Software-Ergonomie* mit *Highlighting, assoziativer Farbgebung, Icon-Technik*, transparenter und konsequenter *Systemphilosophie* und Ansätzen zur *kunstlichen Intelligenz*.

Nachdem sich bei Textverarbeitungssystemen Programme wie Word for Windows und WordPerfect unter Windows weitgehend durchgesetzt haben, gewinnen heute Datenbankprogramme zunehmend an Bedeutung, die insbesondere für Textretrieval geeignet sind. Sie werden den Bekanntheitsgrad von herkömmlichen sogenannten Relationalen Datenbanken wie dBase IV, Ingres, Sybase, etc. bei weitem überflügeln. Damit wird automatisch eine bestimmte Qualität der Leistungsfähigkeit und eine akzeptable **objektorientierte Oberfläche** ihre Verbreitung finden. Gewisse Anteile künstlicher Intelligenz sind ebenso vorhersehbar wie die Standards, die sich schon heute aus dem INTERNET heraus ergeben.

Anschrift des Verfassers

Prof. Dr. Walther Umstätter
Humboldt-Universität zu Berlin
Philosophische Fakultät I
Institut für Bibliothekswissenschaft
Clara-Zetkin-Str. 26
10099 Berlin
Tel.: (030) 20315-236; Fax: -242
E-mail: h0228kdm@rz.hu-berlin.de

Diann Rusch-Feja

Workshop 2: "Bewertung der Qualität von Datenbanken"

Resümé

Herr Dr. Ziegler betonte in seinem Vortrag wichtige Aspekte bei der Kriterienbemessung von Recherche-Software. Vor allem waren Kriterien der Leistung angemerkt. Die Recherche-Struktur soll die Daten verwalten und die Thesaurus-Struktur die Daten zugänglich machen. Die Synonymverwaltung spielt hier eine bedeutende Rolle sowie die Qualität der Indexierung und des Erschließungsinstruments. Anzeigemöglichkeiten für die Rechercheergebnisse sowie die Überschaubarkeit der Bildschirm- oder Menüseiten runden diese Kriterien ab. In der Diskussion wurde erwähnt, daß die meisten Datenbanksysteme keine Thesaurus-Verwaltung haben. Bei LARS, auf dem sowohl FIS-Bildung als auch andere Anbieter ihre CD-ROM-Datenbanken aufbauen, beträgt der Aufpreis für eine Thesaurusverwaltung über DM 10.000,-.

Frau Labouvie hat Aspekte der Datenqualität besprochen. Diese beinhalteten: Multiplizität, inhaltliche Zuverlässigkeit, formale Zuverlässigkeit, Konsistenz, Korrektheit, Mehrsprachigkeit etc. Um die Datenqualität zu sichern, bedarf es Indexierungsregeln, automatischer Unterstützung des Indexverfahrens, Regeln für die sogenannten Phrasen, Klassifikationsregeln, Definition einer Sachgebietsklassifikation etc. Alle 3 - 4 Jahre sollen terminologische Veränderungen geprüft werden. Es gehört zu den Vorgaben, daß ersichtlich ist, nach welchen Gesichtspunkten ein Dokument indexiert wurde. Kurzannotationen erheben den Qualitätswert der Angaben. Scope-Notes zu den Klassifikationen erleichtern den Überblick und ermöglichen Präzision. Mehrsprachige Abstracts tragen zur Internationalität der Datenbank bei, ebenso ein zwei- oder mehrsprachiger Thesaurus. Eine automatisierte Fehlerkontrolle kann in den meisten Datenbanken mittlerweile erfolgreich eingesetzt werden. Schulung der Auswerter und Indexierer ist wichtg, vor allem wo das Autopsie-Prinzip verfolgt wird. Das Datenbankdesign spielt ebenso eine bedeutsame Rolle. Auch auf die Unterschiede in den Datenbanken wurde eingegangen, die durch Vermittlung einer Datenbank in verschiedenen Online-Hosts,

als CD-ROM oder im Rahmen verschiedener CD-ROM-Datenbanken durch die Retrievalsoftware verschiedener Firmen entstehen.

Herr Umstätter behandelte Probleme der Benutzeroberfläche und Qualität der Inhalte. Bisher gab es keine Standards für die Nutzerinterface. Eine stringente Philosophie für die Entwicklung von der Nutzeroberfläche ist Konsistenz. Rückgriffe ergeben mehr Informationen. Hypertext erlaubt Sprünge, die sonst über hierarchische Menüpunkte nicht so direkt sind. Eine "Common Command Language" und die Akzeptanz von Z 39.50 trägt dazu bei, eine einheitlichere und benutzerfreundlichere Oberfläche bei der Mehrzahl angebotener Informationssammlungen zu sichern. Sowohl die Zweisprachigkeit des Retrievalsystems als auch die Möglichkeit, auf zwei Ebenen zu recherchieren ("Endnutzer" und "Experte"), hat Herr Umstätter empfohlen. "Knowledge Finder" eignet sich als Nutzerrecherchetool, zumal es eine neue Art von Thesaurus führt, in dem Synonyme vom System vorgeschlagen werden: "das System erzählt, was man damit machen kann." Auch die Volltextsuche wird in zunehmendem Maße wichtig.

Die Diskussion zeigt zwei Problembereiche auf: Coverage und die Indexierung mittels eines Thesaurus. Coverage hängt hauptsächlich von den Zielen und Vorgaben der Datenbank ab (was wird in der Datenbank ausgewertet und wie) sowie von den vom Hersteller definierten Zugriffspunkten. Betont wurde, daß die Qualität einer Datenbank für ihre Nutzer in erster Linie von der Qualifikation der Indexierer, ihrer Ausbildung und ebenso von den Indexierungsregeln abhängt. Deskriptoren müssen einheitlich und präzise vergeben werden. Bei Kooperationsvorhaben - wie die CD-Bildung von FIS-Bildung mit ca. 20 Zulieferern oder die Datenbank Berufliche Bildung mit ca. 10 Zulieferern - ist die Kompetenz der Auswerter nicht allein maßgebend; jede Institution hat ein ganz individuelles Modell der Schwerpunktsetzung, somit entstehen große Unterschiede bei der inhaltlichen Erschließung. Dasselbe Dokument, das die eine Institution grob verschlagwortet, weil es für die inhaltlichen Schwerpunkte in dieser Institution eher einen Peripheriebereich vertritt, wird bei einer anderen Institution intensiv erschlossen oder unter einem ganz anderen Gesichtspunkt. Datenbanken wie bei der ZPID haben ein eingeübtes Team von Indexierern und eine durchgehende Endredaktion. Trotzdem sind Indexierer-Schulungen und Gegenprüfungen nötig. Bei der FIS-Bildung-Datenbank werden erst im Jahre 1995 Klassifikationsmerkmale für jedes aufgenommene Dokument nachgewiesen. Die Klassifikation muß jedoch den fachlichen Strukturen entsprechen und mehrere Grade der Differenzierung des Themas ermöglichen.

Von einem semantischen Thesaurus war positiv die Rede. Eine Abgleichung von Synonymen muß auf jeden Fall möglich sein. Jeder Endnutzer bringt seinen eigenen Wortschatz mit und denkt in dieser Terminologie. Daher ist eine Synonymverwaltung enorm wichtig, und ein System soll automatisch Synonyme und Relationen unter den Begriffen erkennen können. Automatische

Indexierung oder vom System unterstützte Indexierung soll mit minimalen Zugriffsverlusten möglich sein. Vom Autor definierte sinntragende Wörter im Titel können automatisch erfaßt werden. Formale Aspekte (Autorennamen etc.) können ebenso in der Datenbank auf weiteres Vorkommen geprüft werden, auch im Zusammenhang mit einem bisher verwendeten Schlagwort. Bestimmte Bias sollen möglichst ausgemerzt werden: Der Computer macht den Fehler immer gleich: somit können systematische Fehler in die Datenbank eingebaut werden, die immer schwieriger aufzufinden sind. Gerade bei einer automatischen Indexierung ist deshalb Vorsicht geboten, und die Regel heißt, nichts ohne Prüfung übernehmen. Vor allem gehen mit einer "Übernahme" vorher vergebener Begriffe oft die neuen Begriffe unter.

Mit dem Zuwachs an elektronischen Texten kommt der Volltextrecherche eine neue Bedeutung zu. Trotzdem wird nach wie vor eine Indexierung bzw. inhaltliche Erschließung gebraucht. Im Hinblick auf die neuen elektronischen Medien und Möglichkeiten wurde auch darauf hingewiesen, daß bei der Migration gespeicherter Texte und Informationsquellen auf den neuen technologischen Stand auch Probleme in der späteren Nutzung auftreten. Technische Standards müssen eingehalten werden, sonst kann man eine alte Datenbank, CD-ROM, andere archivierte Quellen etc. nicht mehr lesen. Diesen Archivierungsaufgaben ist in den Einrichtungen entsprechende Bedeutung beizumessen.

Es wurde mehrfach darauf orientiert, bei allen Datenbankfragen die Benutzer zu berücksichtigen und die Schulen und Studenten dabei einzubeziehen. Lösungen für die Qualitätssicherung - sowohl auf der Ebene der Benutzeroberfläche als auch auf der Datenbankstruktur und -inhalte - müssen bezahlbar sein. Zum Schluß der Diskussion gab es bei den Teilnehmern des Workshops weitgehende Einigung darüber, daß vor allem die Aspekte der inhaltlichen Erschließung und Thesaurusfragen in weiteren Gesprächen im Rahmen der GIB-Veranstaltungen fortgeführt werden sollten.

Anschrift der Verfasserin:

Dr. Diann Rusch-Feja
Max-Planck-Institut für Bildungsforschung
Lentzeallee 94
D-14195 Berlin
Tel.: (030) 829 95-230; Fax: (030) 824 99 39
email: ruschfeja@mpib-berlin.mpg.d400.de

Astrid Diening

Literaturverwaltungssysteme im Vergleich

In diesem Beitrag werden die Ergebnisse einer vergleichenden Untersuchung von vier bibliographischen Literaturverwaltungssystemen vorgestellt. Die Analyse und Bewertung dieser Literaturverwaltungssysteme erfolgte für die Informationsvermittlungsstelle der Biologisch-Medizinischen-Sektion der Max-Planck-Gesellschaft.

Folgende Literaturverwaltungssysteme wurden unter dem Betriebssystem MS-DOS analysiert und bewertet:

- EndNote Plus (Cherwell Scientific Publishing), Vers. 1.0
- Pro-Cite (Personal Bibliographic Systems), Vers. 2.0
- Reference Manager (Research Information Systems), Vers. 5.05
- VCH Biblio (VCH Verlagsgesellschaft), Vers. 2.2

Alle vier Systeme sind auf einem IBM-kompatiblen PC einzusetzen. EndNote Plus, Pro-Cite und Reference Manager sind auch als Programm für den Apple-Computer erhältlich. Desweiteren sind seit einiger Zeit Windows-Programme von VCH Biblio und Reference Manager auf dem Markt.

1 Was ist ein Literaturverwaltungssystem und wozu kann man es einsetzen?

Wissenschaftliche Informationen, insbesondere Literaturreferenzen, sind zunehmend in elektronischer Form, z. B. Suchergebnisse aus Online- und CD-ROM-Datenbanken, erhältlich. Eine Verwaltung auf elektronischer Basis bietet sich daher geradezu an, insbesondere bei großen Mengen an Literaturdaten.

Bei den analysierten Literaturverwaltungssystemen handelt es sich um Datenbanksysteme, die hauptsächlich zur Verwaltung von bibliographischen Literaturdaten in elektronischer Form auf einem PC eingesetzt werden können.

Eine Literaturreferenz setzt sich aus verschiedenen Feldern wie Autor (AU), Titel (TI), Quelle (SO) zusammen. Neben den bibliographischen Angaben kann sie noch andere inhaltsbeschreibende Informationen wie ein Abstract oder Schlagworte (Keywords) enthalten. In einem Literaturverwaltungssystem sind die Felder einer Literaturreferenz bzw. deren Inhalte einzeln oder in Kombination suchbar.

Neben den grundlegenden Eigenschaften wie Datenaufnahme, Suchen und Wiederauffinden von Literaturstellen und Datenausgabe ist die Erstellung formatierter Referenzlisten, Bibliographien genannt, ein besonderes Kennzeichen bibliographischer Literaturverwaltungssysteme. Die Anwender erfahren eine wertvolle Unterstützung im Prozeß des Schreibens von Veröffentlichungen, da die Literaturzitate nur einmal in das Literaturverwaltungssystem eingegeben werden müssen. Der Ausgabestil bzgl. Abkürzungen, Zeichensetzung und Schrift kann entsprechend den vom System mitgelieferten Zeitschriftenformaten bestimmt werden. Eine Liste von Literaturreferenzen wird dann für die Ausgabe automatisch entsprechend dem gewählten Stil formatiert.

Der Vorteil von Literaturverwaltungssystemen gegenüber anderen Datenbanksystemen, z. B. dBase, liegt darin begründet, daß ein Literaturverwaltungssystem dem Benutzer eine fertige Datenbankstruktur bietet. Man braucht sich wenige Gedanken über die Definition der verschiedenen Datenfelder und Referenztypen machen. Es kann nach der Installation sofort mit der Dateneingabe begonnnen werden.

2 Welche Kriterien wurden angelegt?

Als Grundlage der Analyse und Bewertung der bibliographischen Literaturverwaltungssysteme wurden Untersuchungskriterien aufgestellt, die sich in der Hauptsache auf Anforderungen zur Bewertung von Retrievalsoftware im allgemeinen und speziell von bibliographischen Literaturverwaltungssystemen stützen (vgl. Tenopir u. Lundeen 1988; Sieverts u. Hofstede 1991; Sieverts et al. 1992). Die Untersuchungskriterien sind folgenden Kategorien zugeordnet:

191

Kategorie	Kriterien
allgemeine Systemmerkmale und Einsatzvoraussetzungen	vom System benötigter Speicherplatz (Arbeitsspeicher u. Festplatte), Netzwerkversion, Mehrplatzsystem
Datenbankeigenschaften	mögliche Beschränkungen, z.B. max. Anzahl Datenfelder pro Datensatz, Anzahl u. Art der Dateneingabemasken
Benutzerschnittstelle	Bildschirmlayout, Eingabeformen, Benutzerhilfen wie Handbuch, Onlinehilfe
Dateneingabe	insbes. Möglichkeit zur automatischen Übernahme von Daten, Dublettenkontrolle
Datenpflege	Editiermöglichkeiten d. Datensssätze, der Eingabemasken
Indexierung	Indexieren bestimmter Datenfelder zum Aufbau einer Indexdatei => schnelle Suche, Aufbau einer Schlagwortliste, eines Thesaurus
Retrieval	Art der Suchfrageformulierung, Boolsche Operatoren, Klammerung von Suchbegriffen, Maskierungsmöglichkeiten
Datenausgabe und Bibliographieerstellung	Sortierung nach Inhalt verschiedener Datenfelder, verschiedene Ausgabeformate, Vorhandensein von Zeitschriftformaten, Kompatibilität zu Textverarbeitungsprogrammen

Abb. 1: Untersuchungskategorien

3 Einzelergebnisse

Im folgenden werden die Ergebnisse der Analyse der Literaturverwaltungssysteme zu den einzelnen Untersuchungskategorien in Tabellenform dargestellt.

Datenbankeigenschaften	EndNote Plus	Pro-Cite	Reference Manager	VCH Biblio
max. Anzahl Felder pro Literaturstelle	27	45	8	32
max. Anzahl Literaturstellen pro Literaturdatei	32 000	2 Mill	65 000	1 Mill
max. Anzahl Literaturdateien	nur begrenzt durch Festplatten-speicher	nur begrenzt durch Festplatten-speicher	eine große DB (65000), mehrere kleinere DB (je 200)	nur begrenzt durch Fest-plattenspeicher
Eingabemaske (z.B. f. Artikel Buch, Dia) - vorgegeben / eigene	14 / 1	20 / 6 pro Literaturdatei	6 / keine	3 / 17
Mehrplatzsystem (gleichzeitiger Zugriff auf dieselbe Literaturdatei)	nein	ja	nein	ja

Abb. 2: Einzelergebnisse des Systemvergleichs in der Kategorie Datenbankeigenschaften

Benutzerschnittstelle	EndNote Plus	Pro-Cite	Reference Manager	VCH Biblio
Bildschirmlayout	aufgabenorientiert	hauptsächlich aufgabenorientiert, z T Programmstruktur i. Vordergrund	aufgabenorientiert	hauptsächlich aufgabenorientiert, z T Programmstruktur im Vordergrund
Eingabeform	hauptsächlich menu-orientiert, Maus	menu- u. kommando-orientiert, keine Maus	menu- u. kommando-orientiert, Maus	menu- u. kommando-orientiert, Maus kryptisch
Fehlermeldung	informativ, z T. zu allgemein	z T. mit Problem-lösung, Warnton,	z T. nur numerischer Code, z T. informativ	unspezifisch, Kontextverlust
Onlinehilfe	z T. nicht im Kontext, nicht sehr umfang-reich	kontextsensitiv, aus-führlich, Lesbarkeit er-schwert: zu lange Zeilen	kontextsensitiv, ver-ständlich, gut lesbar	kontextsensitiv, ganze Bildschirmseite, Ver-lust d. Problembe-zugs, z T. sehr knapp
Handbuch	Englisch, verständlich, gut gegliedert, viele Abb	Englisch, sehr um-fangreich, viel Text, zu wenig Abbildungen	Englisch, verständlich, gut gegliedert	Deutsch, wenig ver-ständlich, schlecht gegliedert, Stichwort-verzeichnis unzurei-chend, wenige Abb

Abb. 3: Einzelergebnisse des Systemvergleichs in der Kategorie Benutzerschnittstelle

automatischer Datenimport	EndNote Plus	Pro-Cite	Referrence Manager	VCH Biblio
Importfunktion in Basisversion	zwei Formate (kommabegrenzt, Feldbezeichner)	ein Format (kommabegrenzt)	keine	keine
Zusatzprogramm	EndLink, Host u. DB unabhängig, Erkennung best. Feldbezeichner	BIBLIO-LINKS, pro Host, kein DIMDI	CAPTURE, pro Host u. DB, über 100 Importmodule	IMPORT, pro Host u. DB, einige Import-module
– modifizierbar	nein	ja	nein	teilweise
Anwendung	einfach	umständlich, aus Pro-Cite nicht aufrufbar	einfach	umständlich, aus VCH Biblio nicht aufrufbar
Dublettenkontrolle	nach Import	nach Import	während des Import-vorganges	nach Import

Abb. 4: Einzelergebnisse des Systemvergleichs in der Kategorie Dateneingabe

BEST COPY AVAILABLE

Datenpflege	EndNote Plus	Pro-Cite	Reference Manager	VCH Biblio
einzelne Literaturstellen	ja	ja	ja	ja
einzelne Literaturdatei				
- Felder umbenennen / löschen / zufügen	einfach	schwierig	nein	schwierig
- Begriffe suchen und ersetzen	nein	ja	ja	ja
einzelne Literaturdatei umbenennen / löschen	DOS-Ebene	ja	nein / ja	DOS-Ebene
einsehbares Register für	Zeitschriftentitel	Titel, Autor, Keywords u.a.	Autor, Keywords, Zeitschriftentitel, Herausgeber	Felder festzulegen

Abb. 5 Einzelergebnisse des Systemvergleichs in der Kategorie Datenpflege

Retrieval	EndNote Plus	Pro-Cite	Reference Manager	VCH Biblio
Suchmodus				
- Standard, d.h. sequentiell alle Lit. stellen der gesamten DB	alle Felder	alle Felder	nur bestimmte Felder Titel, Notes, Dokumentart	alle Felder
- Indexsuche (schnelle Suche)	möglich, einmalige Gesamtindexierung für alle Felder oder nur Feld Keywords notwendig	für Feld Autor, Titel, Keywords möglich, weitere Felder wählbar	für Felder Autor Keywords, Jahr, Hrsg. Refnr., Zeitschriftentitel möglich, Indexierung automatisch	möglich für frei wählbare Felder, einmalige Gesamtindexierung notwendig
Verknüpfungen				
- Boolsche	ja	ja	ja	ja
- relational (von...bis)	nein	ja	ja	ja
Maskierung				
- Standardsuche	Zeichenfolge	rechts (+) links	Zeichenfolge in Feld Notes, Titel, kein Einzelbegriff in d. Feldern	Zeichenfolge, intern, kein Einzelbegriff
- Indexsuche	rechts (autom.)	rechts (+) links	rechts	rechts (+) links
Ansicht des Suchergebnisses	einfach	einfach	einfach	umständlich
- Übersichtsliste	ja	nein, nur Einzelansicht	ja	nein, nur Einzelansicht

Abb. 6: Einzelergebnisse des Systemvergleichs in der Kategorie Retrieval

Datenausgabe	EndNote Plus	Pro-Cite	Referrence Manager	VCH Biblio
als Referenzliste				
– allein	einfach	einfach	einfach	einfach
– in Verbindung mit Manuskript	einfach, TSR-Programm (*)	umständlich	einfach, TSR-Programm (*)	umständlich
kompatibel zu Textverarbeitungs-programm	WinWord, WordPerfect, Word	WordPerfect, Word	WinWord, Word, WordPerfect Win, WordStar 2000, Mass MultiMate	Word, WordStar WordPerfect, TeX
entsprechend Zeitschriften-formaten				
– vorgegeben	10	30	über 100	11
– ändern / neu	umständlich	kompliziert	einfach	kompliziert
Druckfunktion	nein	ja	ja	ja
Austauschformat	kommabegrenzt, Feldbezeichner	kommabegrenzt	Feldbezeichner	kommabegrenzt, Feldbezeichner

Abb. 7 Einzelergebnisse des Systemvergleichs in der Kategorie Datenausgabe.

(*) TSR = Terminate and stay resistant - Programm (speicherresistentes Hilfsprogramm)

Vertrieb	Software	Listenpreis DM (netto) (*)
CHEM Research GmbH Beethovenstr. 7b 60325 Frankfurt Tel. 069 / 74 111 14 Fax 069 / 74 04 96	EndNote Plus (LVS)** EndLink (Importprogramm)	550,- 299,-
Personal Bibliographic Software Inc., fur Europa: Woodside, Hinksey Hill, Oxford OX 5AU, England, Tel. 0044 / 865 326 612, Fax 0044 / 85565 736 354	Pro-Cite (LVS)** Biblio-Link (Importprogramm)	871,- 429,-
Bilaney Consultants GmbH, Schirmerstr. 23, 40211 Dusseldorf, Tel. 0211 7 36 40 43, Fax 0211 7 164 91 77	Reference Manager (LVS)**, Reference Manager Professional complete (LVS + Importmodule)	565,- 895,-
VCH Verlagsgesellschaft mbH, 69469 Weinheim, Tel. 06201 7 60 60	VCH Biblio (LVS)** VCH Biblio Import (Importprogramm)	1450,- 90,-
(*) Stand: Mitte 93, **LVS = Literaturverwaltungsprogramm (Basisversion)		

Abb. 8: Produktangaben

4 Zusammenfassung

Jedes der analysierten Systeme bietet die grundlegenden Funktionen wie Dateneingabe, Datenpflege, Indexierung, Retrieval und Datenausgabe mit Bibliographieerstellung. Bezüglich dieser Grundfunktionen bestehen allerdings Unterschiede in der Ausführung und im Umfang. Das beste Literaturverwaltungssystem gibt es nicht. Jedes der analysierten Systeme weist Stärken und Schwächen auf.

Die Systeme EndNote Plus und Reference Manager überzeugen durch die Gestaltung ihrer Benutzerschnittstelle. Dies ist insbesondere auf die aufgabengerechte Gestaltung des Bildschirmlayouts und eine überwiegend menüorientierte Eingabeform zurückzuführen. Beide Systeme stellen die vom Benutzer zu lösende Aufgabe, nicht die Programmstruktur in den Vordergrund. Dem Einsatz dieser Systeme ist daher gegenüber dem Einsatz von Pro-Cite und VCH Biblio der Vorzug zu geben. Trotz der größeren Flexibilität von EndNote Plus wird Reference Manager als das geeignetere Literaturverwaltungssystem angesehen, da dieses System für wichtige Datenfelder, wie z. B.

Schlagworte, einsehbare Register führt. Diese spielen für die Einhaltung der Begriffskonsistenz bei Eingabe und Suche eine wichtige Rolle.

Pro-Cite und VCH Biblio sind im Vergleich zu EndNote Plus und Reference Manger in einigen Funktionen mächtiger, die Komplexität dieser Systeme erfordert aber eine längere Einarbeitungszeit. Auffallend ist, daß bei diesen Systemen öfters die Programmstruktur im Vordergrund steht. Pro-Cite eignet sich für erfahrene EDV-Nutzer. Der Einsatz von VCH Biblio kann nicht empfohlen werden, da das Programm umständlich in der Anwendung ist und seine Benutzerhilfen unzureichend sind. Der Vorteil der größeren Flexibilität dieses Systems wird dadurch in den Hintergrund gedrängt. Allerdings scheint das Windows-Programm eine Verbesserung zu versprechen.

Literatur:

Tenopir C., Lundeen G. (1988): Managing your information. Neal-Schuman Publishers, New York.

Sieverts EG., Hofstede M. (1991): Software for information storage and retrieval tested evaluated and compared. Part 1: General introduction. *The electronic library 9*, 145-154

Sieverts EG., Figdor J., Bakker S., Hofstede M. (1992): Software for information storage and retrieval tested evaluated and compared. Part 3: End-user software. *The electronic library 10*, 5-19

Anschrift der Verfasserin

Astrid Diening
Diplom-Informationswissenschaftlerin
Bräuhausgasse 5
82205 Gilching
Tel.: (0 81 05) 2 62 43

Willi Reinicke

Zur Auswahl von Softwareprodukten für das Bibliotheks- und das Informationswesen

1 Allgemeines

Dieser Beitrag betrachtet die Problematik der Soft- und HardwareAuswahl und alle damit zusammenhängenden Probleme aus der Sicht einer betrieblichen Einrichtung, in der Aufgaben der Bibliothek und der Information gleichermaßen erfüllt werden müssen (z. B. in Ämtern, industriellen Einrichtungen, Forschungseinrichtungen). Aus der Sicht einer öffentlichen Bibliothek beispielsweise kann sich das eine oder andere Problem etwas anders darstellen. Der Beitrag soll Auswahlkriterien und zu beachtende Randbedingungen benennen.

Die Auswahl eines Programmsystems für einen bestimmten Aufgabenbereich ist eine komplexe Entscheidung. Auf der einen Seite stehen die Aufgaben, die mit Hilfe der EDV automatisiert abgewickelt werden sollen, auf der anderen Seite das große Angebot an Hard- und Software. Dieses ist einer rasanten Entwicklung unterworfen.

2 Charakterisierung des technischen Standes

2.1 Hardware

Auf der Hardwareseite kann man den technischen Stand folgendermaßen umreißen:

- Prozessoren mit einer Verarbeitungsbreite von 16 oder 32 bit bestimmen die Szene,
- Hauptspeicherkapazitäten von 4 MByte sind inzwischen Standard, für anspruchsvolle Programme stehen Hauptspeicherkapazitäten von 8 bis 32 Mbyte zur Verfügung (über 4 Mbyte oft durch eine nachträgliche Hauptspeichererweiterung, die Preisfrage ist hier zu berücksichtigen),

- Geräte mit Festplattenspeicherkapazitäten von 300 MByte aufwärts bis in den GByte-Bereich (im GByte-Bereich z. B. als Server in Netzen) sind im Handel,
- CD-ROMs sind auf einem raschen Vormarsch und als Speichermedium (Software und Datenbasen werden in einem immer größeren Umfang auf CD-ROM angeboten) inzwischen unverzichtbar,
- die PC-Technik öffnet sich für multimediale Anwendungen (Sound, Animationen, Videos).

2.2 Software

Softwaremäßig läßt sich die Entwicklung allgemein so charakterisieren:

- Bei den Betriebssystemen wird der Zeitabstand des Auftretens neuer Betriebssystemversionen immer geringer,
- das WINDOWS-Betriebssystem und das Betriebssystem OS/2 drängen auf den Markt (das Betriebssystem OS/2 nach langem Anlauf),
- UNIX wird vor allem dort benutzt, wo es auf einen Multiuser-Betrieb ankommt,
- das Betriebssystem MS-DOS ist trotzdem noch am weitesten verbereitet (z. B. bei der CD-ROM-Anwendung in Deutschland, Österreich und der Schweiz 89 %, in den USA 73 %),[1] in Bezug auf die Funktionalität und Bedienung der Programmsysteme,
- der Vormarsch der Systeme mit graphischen Bedienoberflächen,
- das Angebot multimedialer Software,
- die Netzwerkfähigkeit der Software,
- die Zunahme der Funktionalität: Bei der Software für den Bibliotheks- und Informationsbereich läßt sich diese Tendenz sehr gut beobachten. Anfangs war die Situation von Programmen, die einzeln auf die Informations- und Dokumentationstätigkeit oder das Bibliothekswesen zugeschnitten waren, geprägt. Heutzutage sind neben den Programmen für Einzelprobleme zunehmend Programme mit einem komplexen Funktionsangebot für beide Bereiche zu verzeichnen, mit denen es z. B. möglich ist, den gesamten Geschäftsgang einer Bibliothek einschließlich der Aufgaben der Informationsversorgung abzudecken.
- Auf dem Gebiet der sogenannten Standardsoftware und auch bei der Spezialsoftware werden immer mehr Produkte angeboten. Neue Produkte reten auf, vorhandene Produkte werden weiterentwickelt.

[1] nach de Kemp; Junge: Wissenschaftliche Zeitschriften auf CD-ROM, *Bibliotheksdienst* 28 (1994) 10, S. 1651-1660

Das Softwareangebot auf dem hier zur Diskussion stehenden Gebiet ist beim Berliner Arbeitskreis für Information in einer Datenbank "SOFTWA" mit der in solchen Fällen möglichen Vollständigkeit erfaßt. Diese Datenbank umfaßt z. Z. 844 Produkte und gibt wertungsfrei Informationen zu den Produzenten/Anbietern, den technischen Einsatzbedingungen, den Funktionen und Leistungen, Einsatzfällen und sonstigen Parametern. Sie steht für Auskünfte über Frau Pott[2] zur Verfügung.

Die Datenbank beinhaltet folgende Programmarten:
- Bürosyteme,
- Datenbanksysteme für optische Speichermedien (CD-ROM),
- Dokumentenlieferung,
- Scannen,Texterkennen,
- Enduser-/Bibliographie-/Persönliche Informationssysteme,
- Wörterbuch-/Thesaurussysteme,
- Expertensysteme,
- Integrierte Bibliothekssysteme,
- Multimediasysteme,
- Datenkommunikation/Online-Recherchen,
- Technische Dokumentation,
- Standardsoftware,
- Literaturdokumentation,
- Bibliothekarische Einzelfunktionen,
- Dokumentarische Einzelfunktionen,
- Archivfunktionen.

2.3 Kommunikationstechnologien

Die Entwicklung im Informations- und Bibliothekswesen wird wesentlich durch die rasante Entwicklung auf dem Gebiete der Kommunikationstechnologien bestimmt. Entwicklungen wie "data highway" beeinflussen die Informations- und Bibliotheksszene beträchtlich. Über internationale Datennetze hat der Informationssuchende, z. B. über das Internet, Zugriff zu Bibliothekskatalogen (OPAC) oder zu Datenbanken von kommerziellen Anbietern. Dadurch hat der Nutzer Zugang zu solchen großen Ressourcen wie der British Library, der Library of Congress, großen Universitätsbibliotheken und anderen leistungsfähigen Einrichtungen. Die Herkunft der Informationen ist in der Regel für den Nutzer sekundär, für ihn ist es wichtig, die benötigten Informa-

[2] BAK Information, c/o TU Berlin, Geschäftsstelle Gebäude FRA-B, Franklinstr. 15, 10587 Berlin; Tel.: (030) 314-24741,-22217; Fax: (030) 314-24743; E-mail: pottswel@mailszrz.zrz.tu-berlin

tionen zu erhalten, egal, ob sie einer Bibliothek entstammen oder einem Bereich, den man der "Information und Dokumentation" zurechnet.

3 Softwareauswahl für den Bibliotheks- und Informationsbereich

3.1 Allgemeine Anforderungen

Wenn es um die Einführung der rechnergestützten Arbeit in Prozessen der Information und Dokumentation und in der Bibliothek geht, dann besteht die Aufgabe, die geeignete Hard- und Software, auch unter zukunftssicheren Aspekten, auszuwählen.

Wie soll nun vorgegangen werden, welche Kriterien sollen für den Softwareeinsatz zugrundegelegt werden?

Welche Anforderungen stellt allgemein ein Nutzer aus der Sicht des Zusammenwirkens von Bibliothek und Information?

- Er möchte in der Regel möglichst viele Arbeitsgänge rechnergestützt abarbeiten und dabei die einmal gespeicherten Informationen optimal nutzen.
- Er möchte eigene (interne) Informationen speichern, recherchieren und verarbeiten können.
- (Im Bedarfsfall muß auch eine Zulieferung zu Datenbasen von Hosts möglich sein).
- Er möchte fremde Informationen (unter Wahrung des Urheberrechtes selbstverständlich) aus über Hosts angebotenen Datenbasen, von CD-ROM oder aus dem Internet für seine Aufgaben auf der lokalen Ebene speichern, recherchieren und weiterverarbeiten können. (Diese Dateien werden allgemein als ASCII-Files angeboten).
- Er möchte die Gewähr haben, daß gültige Normen für die bibliographische Beschreibung und die inhaltliche Erschließung oder für den Datenaustausch eingehalten werden, so daß aus der rechnergestützten Verarbeitung stammende Informationen normgerecht (z. B. DIN 1502, 1505) in Leistungen wie Bibliographien, Literaturverzeichnissen usw. eingehen.

3.2 Kriterien für den Auswahlprozeß

Anhand welcher Kriterien sollte der potentielle Anwender, gemessen an seinem Arbeitsprozeß, sein Programmsystem auswählen? Es soll noch einmal darauf verwiesen werden, daß auch diese Kriterien aus der Ganzheitsbetrach-

tung von Bibliotheks- und Informationswesen zu sehen sind. Bei dem Auswahlprozeß sollten die folgenden Komplexe berücksichtigt werden:

- Funktionen und Leistungen,
- Systemaufbau,
- technische Voraussetzungen,
- Bedienung, Nutzerfreundlichkeit,
- Kundenservice,
- Preis,
- Referenzen.

Die einzelnen Komplexe sollen etwas näher betrachtet werden.

3.2.1 Funktionen und Leistungen

3.2.1.1 Arbeitsgänge

Die Forderung, die Arbeitsgänge in der Bibliotheks- und Informationsarbeit möglichst komplex zu unterstützen, bedingt die Ankopplung an diese Arbeitsgänge. Charakteristische Arbeitsgänge in dieser technologischen Kette sind:

- die Ermittlung der relevanten Informationsquellen zur Beschaffung (aus dem Vorankündigungsdienst, aus Angebotslisten der Verlage, aus Bibliographien (in gedruckter Form, als Disketten- oder CD-ROM-Dienst) und Auslösung des Bestellvorganges (durch Kauf, Tausch),
- bei dem Eingang der Quelle Einarbeitung, d. h. bibliographische Beschreibung und inhaltliche Erschließung, Speicherung,
- Ausleihe: Ausleihverbuchung und -mahnung,
- Austausch mit anderen Systemen und Einrichtungen,
- Suche und Aufbereitung relevanter Informationen (Bibliographie, SDI, Recherchen sollen hier genannt werden).

Darüber hinaus sind noch zusätzliche Fragen von Interesse. Wie unterstützt ein System den Aufbau und die Pflege von Dokumentationssprachen? Wie unterstützt ein Programmsystem insgesamt diese Prozesse? Wie sind die Verarbeitungsgeschwindigkeiten bei diesen Prozessen?

3.2.1.2 Output und Input

Output und Input eines Programmsystems sind maßgebliche Faktoren für die Nutzung eines Programmsystems.

Bezüglich des **Outputs** sind von Interesse:

- Möglichkeit der DIN- bzw. RAK-gerechten Literaturaufbereitung,
- Unterstützung auch noch vorhandener manueller Arbeitsweisen, z. B. durch die Möglichkeit der Bibliographie-, Listen- und Titelkartenausgabe,
- Recherchemöglichkeiten: feldbezogene Suche und Volltextrecherche (u. a. mit Trunkierung der Suchbegriffe),
- Einblendung von Wort- bzw. Indexlisten (beispielsweise aus Dokumentationssprachen oder auch selbst generierte) bei der Suche,
- Suchmethodik: sequentielle und Indexsuche, daraus resultierende Recherchegeschwindigkeit,
- Möglichkeit der Speicherung von Suchfragen bzw. Suchprofilen, um solche Leistungen wie z. B. SDI erzeugen zu können,
- Zusammenarbeit mit anderen Systemen, d. h. Übergabe von Nachweisen an andere Systeme, Datenaustausch und Datentransfer (inhaltliche Transformation und rechentechnische Anpassung),
- Möglichkeit der leichten Generierung eigener Aufbereitungsformen,
- Möglichkeit der layoutmäßigen Weiterverarbeitung durch Textverarbeitungs- bzw. DTP-Programme (z. B. Winword. AmiPro, Pagemaker).

Bezogen auf den **Input** sollten die folgenden Kriterien erfüllt werden:

- Möglichkeit der einfachen Generierung eigener Eingabeformate und Eingabemasken,
- quantitative Grenzen in den Eingabeformaten,
- entsprechend dem Bedarf Eingabemöglichkeit von Literaturinformationen, Fakteninformationen, Volltexten (auch mit graphischen Informationen),
- verwendbare Eingabedatenträger,
- einfache Korrekturmöglichkeiten,
- Möglichkeit der Dublettenkontrolle bei der Eingabe,
- Einblendung von Wort- bzw. Indexlisten in die Erfassung,
- syntaktische und semantische Prüfungen bei der Eingabe,
- Zusammenarbeit mit anderen Systemen, d. h. Übernahme von Nachweisen aus anderen Systemen.

3.2.2 Systemaufbau

Auf dem Markt werden Programmsysteme angeboten, die die gesamte Arbeitsbreite in der Information und Bibliothek abdecken (natürlich auch mit

dem entsprechenden Preis: Preise von über 10 000 DM sind dabei möglich). Größere Einrichtungen streben oft nach solchen Lösungen. Kleinere Einrichtungen ohne diese Aufgabenbreite geben sich mit einfacheren, nicht so komplexen Lösungen zufrieden, die an ihr Aufgabenspektrum angepaßt sind.

Ideal für beide Einsatzfälle sind Programmsysteme mit einem modularen Aufbau, bei denen es möglich ist, auch einzelne Systemkomponenten zu erwerben (natürlich dann auch mit einer angepaßten Preisgestaltung) und auch separat zu betreiben.

3.2.3 Technische Voraussetzungen

Mit der Softwareauswahl sind immer die Entscheidungen verbunden

- über das zu benutzende Betriebssystem,
- zwischen der Nutzung von auf dem Markt angebotener Standardsoftware oder der Schaffung einer eigenen Lösung.

Betriebssystem

Da Datenverarbeitungsprojekte für die Information und für die Bibliothek im unternehmerischen Rahmen nicht als isolierte Lösungen bestehen, sondern in das Organisationsregime des Unternehmens eingegliedert sind, wird die Betriebssystementscheidung dem Anwender oft durch eine im betrieblichen Rahmen gefällte Entscheidung abgenommen. Ist das nicht der Fall, sollte der potentielle Anwender bei der Betriebssystemwahl folgende Aspekte berücksichtigen:

- DOS und das Windows-Betriebssystem stellen durch die annoncierte Programmkompatibilität und die Vielfalt verfügbarer Anwenderprogramme, auch unter dem Aspekt der Netzwerkarbeit, risikolose Lösungen dar.
- Wo unter Multiuserbedingungen auch mit anderen Systemen zusammengearbeitet werden soll, ist das UNIX-System in die Diskussion einzubeziehen.
- Zum Betriebssystem OS/2 liegen beim Autor keine Erfahrungen vor.

Hardwareanforderungen

Die Mindesthardwareanforderungen werden von den Programmsystemanbietern genannt. Sie sind mit der zur Verfügung stehenden Techik zu vergleichen, und dann ist die notwendige Entscheidung zu fällen. Ein besonderes Augenmerk sollte der Graphikkarte und dem Monitor gewidmet werden, insbesondere wenn es um die Speicherung von Volltexten geht.

Bei der Entscheidung über die Nutzung einer auf dem Markt von professionellen Anbietern erstellten Lösung oder der Schaffung einer eigenen Lösung sind zwei Faktoren zu berücksichtigen:

- die notwendige EDV-mäßige Qualifikation, gekoppelt mit soliden Kenntnissen im Bibliotheks- und Informationswesen, als Voraussetzung für die Schaffung einer eigenen Lösung,
- eigene Lösungen sind meist isolierte Lösungen und führen oft bei den Fragen eines Informationsaustausches und der Einhaltung der Standards zu Problemen.

3.2.4 Bedienung, Nutzerfreundlichkeit

Zur Beurteilung der Bedien- und Nutzerfreundlichkeit sollten folgende Gesichtspunkte beachtet werden:

- Einfache Systeminstallation,
- Art und Übersichtlichkeit der Bedienoberfläche,
- Umgang mit dem System, Benutzerführung, Bedienkomplexität (Komplexität der zur Erreichung eines Zieles notwendigen Bedienhandlungen), Selbsterklärungsfähigkeit, Verläßlichkeit, Eindeutigkeit und Transparenz der Dialogführung, Aufwand bei Fehlerkorrekturen, Möglichkeit des Rückgängigmachens von Aktionen, Systemkonsistenz, Übersichtlichkeit und Einheitlichkeit beim Bildschirmaufbau,
- Unterstützung des Nutzers bei Fehlern,
- Gestaltung des Anfangsmenüs, Programmeinstieg,
- Gestaltung der "Hilfe"-Funktionen,
- Aufgabenangemessenheit, d. h. wie unterstützt das System den Anwender bei der Aufgabenbearbeitung, ohne ihn mit Systemkenntnissen zu überfordern,
- Lernunterstützung,
- Komfort bei der Datenerfassung und Datenkorrektur,
- Qualität der Dokumentation bzw. des Handbuches, insbesondere zur Fehlerbehandlung.

3.2.5 Kundenservice

Zur Beurteilung des Kundenservice können folgende Kriterien herangezogen werden:

- Unterstützung des Lieferanten bei der Sicherstellung der Funktionen,
- Konditionen bei Updates,
- Durchführung von Schulungen (gehören diese zum Leistungsumfang oder sind diese zu bezahlen),
- Serviceleistungen, z. B. Reaktion bei auftretenden Fehlern,
- Lieferbedingungen.

3.2.6 Preis des Gesamtsystems oder einzelner Komponenten

Das Preis- Leistungsverhältnis kann nur durch den Vergleich zwischen den einzelnen potentiellen Produkten ermittelt werden. Preisempfehlungen für bestimmte Funktions- Leistungsumfänge sollen und können an dieser Stelle nicht gemacht werden. Es sei aber auf die Datenbank beim Berliner Arbeitskreis für Information verwiesen.

3.2.7 Referenzen

Wie zuverlässig ist der Lieferant? Welche Referenzen gibt es für das Softwaresystem? Welche Aussagefähigkeit haben die Referenzen? Die Frage der Referenzen ist deshalb wichtig, weil mit einer Vielzahl von Anwendungspartnern auch die Gewähr für eine gewisse Leistungsfähigkeit des Systems gegeben ist und für Problemfälle Konsultationspartner vorhanden sind.

Anschrift des Verfassers:

Dr. Willi Reinicke
PROGRIS GmbH Berlin
Löcknitzstr. 45
12587 Berlin
Tel. (030) 82 55 049, Fax: 82 57 646

Bernd Christopher

Workshop 3: "Software-Entscheidungen Kriterien zur Auswahl von Literaturverwaltungsprogrammen"

Zusammenfassung

Der Workshop 3 stellte interessierten Teilnehmern ein Forum zur Verfügung, in dem aus der Sicht von Fachwissenschaftlern, Bibliothekaren und Softwareentwicklern über die entscheidenden Kriterien für eine Softwareentscheidung diskutiert wurde.

Einig waren sich die Teilnehmer darin, daß die Interessen von Softwareentwicklern, Datenbankbetreibern und Datenbankbenutzern nicht in jedem Fall identisch sind. Die Kriterien leichtes Erfassen, sicheres Speichern, unkomplizierte Suche und Ausgabe sind die gemeinsame Ausgangsbasis der verschiedenen Interessengruppen.

Grundlage für die Diskussion waren vorbereitete Statements, die aus der unterschiedlichen Sicht von Fachwissenschaftler, Bibliothekar und EDV-Spezialist einen Einblick in die Problematik gaben.

Astrid Diening (Gilching) hat in einer Untersuchung vier Literaturverwaltungssysteme aus Sicht einer Fachwissenschaftlerin gegenübergestellt und anhand von Kriterien (z. B. Dateneingabe, Bildschimlayout, Indexierung, Retrieval, Datenausgabe, Benutzerschnittstellen) analysiert. Fazit dieser Analyse war, daß die Bedürfnisse des Nutzers / Anwenders ein wesentlicher Punkt sind, da keines der Programme in allen Bereichen ausreichend positiv abgeschnitten hat.

Dr. Karl-Heinz Sühnhold (Berlin) ging auf einige aus Sicht der Bibliothekare wichtige Kriterien ein, denen eine anspruchsvolle Bibliothekssoftware genügen muß. Die Berücksichtigung bibliothekarischer Regeln, Fremddatenübernahme, Netzwerkfähigkeit, Austauschfähigkeit zwischen den Formaten, Anlage von Stammsätzen, Bedienbarkeit und individuelle Anpassungen, OPAC mit Standortnachweis, unterschiedliche Recherchewege, Dublettenkontrolle ... sind Aspekte, die eine Bibliothek entsprechend ihren Bedürfnissen gewichten muß. Die kurze Auswertung einer Befragungsaktion des Deutschen Bibliotheksinstituts zum DV-Einsatz und verwendeter Software in deutschen Bibliotheken rundete den Vortrag ab.

Dr. Willi Reinecke (Berlin) sprach über einige Kriterien, die Hard- und Software aus Sicht des EDV-Spezialisten erfüllen sollten. Das Kriterium "Was will ich erreichen?" muß im Mittelpunkt stehen. Da das Angebot an Software unüberschaubar ist, ist es erforderlich, die Kriterien für die Softwareentscheidung an den eigenen Bedürfnissen (auch perspektivisch) zu orientieren.

Eine problemlose Installation, automatische Systemanpassungen, übersichtliche Oberflächen, Gestaltung von Ein- und Ausgabemasken, hohe Verläßlichkeit und Transparenz, ein gutes Handbuch, Kundenservice, Preis-Leistungs-Verhältnis ... sind aber Aspekte, die immer mit in eine Entscheidung einbezogen werden sollten.

Bei den Beiträgen und Diskussionen wurde deutlich, daß die Kriterien hinsichtlich Aufbau und Benutzung von Datenbanken unterschiedlichen Stellenwert haben.

Eine Software, die den Ansprüchen von Datenbankbetreibern und -benutzern in jeder Hinsicht gerecht wird, ist bisher nicht zum Einsatz gekommen. Zu unterschiedlich sind die Wünsche und zu groß das vorhandene Angebot. Hier gilt es, den begonnenen Dialog zwischen den Softwareentwicklern, Datenbankbetreibern und den Endnutzern fortzusetzen.

Die Gesellschaft Information Bildung sollte diese Diskussion auch in Zukunft aufgreifen und den Dialog zwischen den Interessengruppen fortführen.

Anschrift des Verfassers

Dr. Bernd Christopher
Bundesinstitut für Berufsbildung
Fehrbelliner Platz 3
10707 Berlin
Tel.: (030) 8643-2230, Fax: -2607

Teilnehmerverzeichnis

Bahnsen, Christian, Universität Hamburg, Fachbereich Erziehungswissenschaft
Bambey, Doris, Modellversuch FIS Bildung, Eschborn
Bauer, Dr. Sabine, Berliner Institut für Lehrerfort- und -weiterbildung und Schulentwicklung (BIL)
Becker, Christine, Universität Potsdam, ZEIK/STN-Schulungszentrum
Bergmann, Harald, Universität Jena, Institut für Erziehungswissenschaften
Bertelsmann, W. Arndt, W. Bertelsmann Verlag GmbH & Co KG, Bielefeld
Bierwagen, Dr. Marion, DIPF, Bibliothek für Bildungsgeschichtliche Forschung, Berlin
Bodem, Claudia, Staats- und Universitätsbibliothek Bremen
Borkenhagen, Frederik, Deutsche Vereinigung für Sportwissenschaft, Universität Hamburg
Botte, Alexander, Modellversuch FIS Bildung, Eschborn
Christopher, Dr. Bernd, Bundesinstitut für Berufsbildung, Berlin
Cyprian, Rüdiger, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg
Dechsling, Dr. Rainer, Verlag C.-H. Beck, München
Delling, M. A., Manfred, Deutsches Institut für Fernstudienforschung, Tübingen
Diening, Astrid, Gilching bei München
Diepold, Prof. Dr. Peter, Humboldt-Universität zu Berlin, Abt. Pädagogik und Informatik
Dietrich, Peter, Hamburg
Eigemeier, Manfred, W. Bertelsmann Verlag GmbH & Co KG, Bielefeld
Feuerstein, Prof. Dr. Thomas, Fachhochschule Wiesbaden
Fingerle, Prof. Dr. Karlheinz, Universität-Gesamthochschule Kassel, Fachgebiet Erziehungswissenschaft
Flemming, Simone, Bundesinstitut für Berufsbildung, Bilbiothek/Dokumentation, Berlin
Fröhling, S., Berliner Institut für Lehrerfort- und -weiterbildung und Schulentwicklung (BIL), Berlin
Funk, Prof. Dr. Robert, Humboldt-Universität zu Berlin, Institut für Bibliothekswissenschaft
Furrer, Max, Pestalozzianum Zürich, Bibliothek/Mediothek

Grebing, Dr. Renate, Universität Marburg, Infozentrum für Fremdsprachenforschung
Günther, Gerda, Pädagogisches Landesinstitut Brandenburg, Bibliothek, Ludwigsfelde
Handlem, Derek, The British Council, Berlin
Hartlaub, Melanie, Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt/M
Hochwald, Dr. Karl-Heinz, Landesinstitut für Schule und Weiterbildung, Soest
Horn, Dipl.-Päd. Klaus-Peter, Humboldt-Universität zu Berlin, Institut für Allgemeine Pädagogik
Jentzsch, Irene, Freie Universität Berlin, ZEAM
Kendzia, Michael, INBAS, Frankfurt/M
Klingmüller, Dr. Bernhard, Freie Universität Berlin, FB Erziehungs-und Unterrichtswissenschaften, Projekt WISTA
von Knobelsdorff, Angelika, Technische Universität Berlin, Universitätsbibliothek
Knost M.A., Peter, Humboldt-Universität zu Berlin, Vergleichende Erziehungswissenschaft
Kohls, Dr. Rosemarie, Projektgruppe Hochschulforschung, Literaturdatenbank Hochschulbildung, Berlin
König, Gerhard, Fachinformationszentrum Karlsruhe
Krause, Ute, Berliner Institut für Lehrerfort- und -weiterbildung und Schulentwicklung (BIL), Bibliothek
Kreusch, Julia, Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt/M.
Labouvie, Yrla M., Zentralstelle für Psychologische Information und Dokumentation, Trier
Liebl, Peter, Deutsches Institut für Erwachsenenbildung, Frankfurt/M
Löhr, Paul, Internationales Zentralinstitut für das Jugend- und Bildungsfernsehen, München
May, Anne, Fachreferat Erziehungswissenschaften, Universitätsbibliothek, Paderborn
Meyer, Dipl.-Inf. Frerk, Humboldt-Universität zu Berlin, Abt. Pädagogik und Informatik
Meyer, Wolfgang, Schulzentrum Ekel, Norden
Ni Cheallaigh, Martina, CEDEFOP, Berlin
Paul, Dr. Peter, Humboldt-Universität zu Berlin, Zweigbibliothek Erziehungswissenschaften, Berlin
Pipial, Rita, Berliner Institut für Lehrerfort- und -weiterbildung und Schulentwicklung (BIL), Berlin
Piwka, Elisabeth, Bundesinstitut für Berufsbildung, Berlin
Prüstel, Sabine, Bundesinstitut für Berufsbildung, Berlin
Reinicke, Dr. Willi, PROGRIS GmbH, Berlin
Reinsch, Ines, Bundesinstitut für Berufsbildung, Berlin

Richter, Dr. Christian, Humboldt-Universität zu Berlin, Abt. Pädagogik und Informatik
Ritzi, Christian, DIPF, Bibliothek für Bildungsgeschichtliche Forschung, Berlin
Rommel, Dr. Hans-Georg, EURYCLEE, Rheinbach
Rosenkranz, Bettina, Volk und Wissen Verlag GmbH, Berlin
Rost, Dr. Friedrich, Freie Universität Berlin, Fachbereich Erziehungs- und Unterrichtswissenschaften
Rusch-Feja, Dr. Diann D., Max-Planck-Institut für Bildungsforschung, Berlin
Scheffler, Dr. Christine, Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt/M
Schneider, Stefan, Hochschule der Künste, Berlin
Schray, Dorothee, Staatsinstitut für Schulpädagogik und Bildungsforschung, München
Schubert, J., Sächsisches Staatsinstitut für Bildung und Schulentwicklung/ Comenius-Institut, Dresden
Sörensen, Bernd, Hochschule der Künste, Berlin
Sonnenburg, Peter, Humboldt-Universität zu Berlin, Abt. Grundschulpädagogik
Springer, Sylvia, Universität Düsseldorf, Fachbereich Erziehungswissenschaften
Stenzel, Brigitte, Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt/M.
Steuber, Dr. Hartmut, Heinrich-Heine-Universität, Erziehungswissenschaftliches Institut, Düsseldorf
Straub, Dipl.-Ing. Werner, Landesinstitut für Schule und Weiterbildung, Soest
Sühnhold, Dr. Karl-Heinz, Deutsches Bibliotheksinstitut, Berlin
Sziburies, Dr. Ute, Martin-Luther-Universität Halle, Zweigbibliothek Erziehungswissenschaften
Theers, Ralf, Hessisches Institut für Bildungsplanung und Schulentwicklung, Wiesbaden
Ullmann, Monika, Technische Universität Chemnitz/Zwickau, Universitätsbibliothek
Umstätter, Prof. Dr. Walther, Humboldt-Universität zu Berlin, Institut für Bibliothekswissenschaft
Wederhake-Gottschalk M. A., Monika, Universitätsbibliothek, Dortmund
Wicker, Katrin, Modellversuch FIS Bildung, Eschborn
Ziegler, Dr. Heinz, Projekt Literaturdatenbank Berufliche Bildung, Berlin